Stabilité des Systèmes à Événements Discrets Stochastiques Approche Algébrique

Jean Mairesse

Juin 1995

[1]Le présent document a été établi en exécution du contrat n° 9181526 passé par la Direction des Recherches Etudes et Techniques, Direction Scientifique / Section Soutien à la Recherche (Délégation générale pour l'armement.)

Je tiens à remercier très sincèrement François Baccelli. Je tiens pour un privilège d'avoir pu effectuer ma thèse sous sa direction. Il a guidé et soutenu mon travail avec une disponibilité maximale et très motivante. Sa compétence et sa curiosité ont été une source constante de questions intrigantes et de réponses éclairantes. Certains des travaux présentés dans cette thèse ont été réalisé en commun avec lui. Tous ont profité de ses remarques et conseils.

Je tiens à exprimer ma plus vive gratitude envers Jacques Neveu. J'ai eu la chance de suivre ses cours à l'Ecole Polytechnique. Par leur clarté et leur rigueur remarquables, ils m'ont fait apparaître comme une évidence qu'il fallait étudier les probabilités. Je tiens ensuite à le remercier de m'avoir encadré avant ma thèse et conseillé au cours de celle-ci. Enfin, je considère comme un honneur qu'il ait accepté de présider mon jury.

Je remercie vivement Venkat Anantharam, Mike Keane et Michel Viot d'avoir accepté de rapporter sur ma thèse. Leurs remarques ont permis d'améliorer le manuscrit mais aussi de mettre en lumière des perspectives de recherches futures. Mes remerciements vont également à Philippe Bougerol et Guy Cohen qui ont bien voulu faire partie de ce jury.

Ce fut un honneur mais aussi un plaisir de profiter des années sabbatiques passées par Alexander Borovkov et Serguei Foss au sein du projet Mistral pour travailler avec eux et profiter de leur expérience. Mon travail a également profité et à de très nombreuses reprises des connaissances encyclopédiques de Stéphane Gaubert. Enfin je tiens à remercier Jean-Marc Vincent pour le séjour intéressant et agréable que j'ai effectué dans son laboratoire.

Un grand merci, collectivement et individuellement, aux membres du projet Mistral, pour l'atmosphère dynamique, chaleureuse et amicale qui y règne. J'ai eu quotidiennement l'occasion d'échanger des idées ou de profiter des conseils de collègues-amis tels que Bruno Gaujal (colocataire presque parfait), Lucian Finta, Miguel Canales, Alain Jean-Marie ou Damien Artiges (entre autres et dans le désordre).

J'espère que le caractère de figure imposée de l'exercice des remerciements n'aura pas masqué la sincérité de mon propos. Encore merci à tous.

Table des matières

0 Introduction **7**
- 0.1 Systèmes à Evénements Discrets . . . 7
- 0.2 Approche algébrique . . . 7
- 0.3 Systèmes Déterministes-Stochastiques . . . 8
- 0.4 Résumé et Contributions . . . 10
- 0.5 Problèmes Non Résolus . . . 12

I Systèmes Linéaires Déterministes **13**

1 Exemples de Systèmes (max,+) Linéaires **14**
- 1.1 Introduction . . . 15
- 1.2 Modèle Mathématique . . . 15
- 1.3 Réseaux de Petri . . . 16
  - 1.3.1 Présentation générale . . . 17
  - 1.3.2 Atelier flexible . . . 21
  - 1.3.3 Files d'attente avec blocage . . . 23
- 1.4 Automates temporisés . . . 26
  - 1.4.1 Un modèle de stockage . . . 28
  - 1.4.2 Réseaux de Petri avec choix . . . 29
- 1.5 Optimisation Dynamique . . . 31
- 1.6 Mécanique Statistique . . . 33
- 1.7 Conclusion . . . 35

2 Minimal Representation of Uniform Recurrence Equations **36**
- 2.1 Introduction . . . 37
- 2.2 Basic Models . . . 38
  - 2.2.1 Dependence graph . . . 38
  - 2.2.2 Reduced graph . . . 39
  - 2.2.3 Recycled case . . . 40
  - 2.2.4 Preliminary remarks . . . 40
- 2.3 Parallel Executions . . . 41
  - 2.3.1 Relations with the basic scheduling problem . . . 41
  - 2.3.2 Pebble game . . . 42
  - 2.3.3 Summary . . . 46
- 2.4 Cuts and Pebbles . . . 48
  - 2.4.1 Cuts in $\mathcal{D}$ . . . 48

2.4.2 Relations with executions of Game 1 . . . 50
2.4.3 Complexity results . . . 52
2.5 Cuts and Delays . . . 53
2.5.1 Retiming . . . 53
2.5.2 Counting the delays . . . 55
2.5.3 Cuts and delays . . . 56
2.5.4 Summary and open problems . . . 60
2.6 Application 1 : Registers in Circuit Design . . . 60
2.6.1 Definition of a circuit . . . 60
2.6.2 Counting the registers . . . 62
2.6.3 Minimizing the registers . . . 62
2.6.4 Complexity results . . . 64
2.7 Application 2: (max,+) Linear Systems and Parallel Simulation . . . 64
2.7.1 Introduction . . . 65
2.7.2 MPR of order 1 . . . 67
2.7.3 Parallel simulation of time varying MPR . . . 67
2.7.4 Optimization results . . . 68
2.7.5 Backward transformation . . . 70
2.7.6 Stochastic issues . . . 71

3 Graphical Approach of the Spectral Theory in the (max,+) Algebra 72
3.1 Introduction . . . 73
3.2 The (max,+) Algebra . . . 73
3.3 A Simple Manufacturing Model . . . 74
3.4 Spectral Theory in $\mathbb{R}_{max}$ . . . 75
3.4.1 General presentation . . . 76
3.4.2 The projective space . . . 78
3.4.3 Change of basis . . . 80
3.4.4 Classification . . . 82
3.5 Illustrated Spectral Theory in Dimension 3 . . . 83
3.5.1 Scs1-Cyc1 . . . 83
3.5.2 Scs3-Cyc1 and Scs1-Cyc3 . . . 84
3.5.3 Scs2-Cyc1 and Scs1-Cyc2 . . . 87
3.5.4 Scs2-Cyc2 . . . 91
3.5.5 Transient regimes . . . 92
3.6 Application to the Manufacturing Model . . . 95
3.7 A Projectively Infinite Semigroup of Matrices . . . 96
3.8 Conclusion . . . 99

4 Application in Cyclic Scheduling 100
4.1 Introduction . . . 101
4.2 Periodicity of the Schedule . . . 103
4.3 Graphical Illustration . . . 105
4.4 Quantitative Results . . . 108

5 Algorithms **110**
5.1 Algorithm . . . 111
5.2 Complexity . . . 113
5.3 Alternative Algorithm . . . 113

II Systèmes Linéaires Stochastiques **116**

6 Products of Random Matrices in the (max,+) Algebra **117**
6.1 Introduction . . . 118
6.2 Two Motivating Models . . . 120
6.2.1 Task graphs . . . 120
6.2.2 Cyclic Jackson network . . . 120
6.3 (max,+) Algebra . . . 123
6.4 Deterministic Spectral Theory . . . 125
6.5 Semigroup of Matrices . . . 128
6.6 Borovkov's Renovating Events Theory . . . 129
6.7 Presentation of the Results . . . 131
6.7.1 Results from Baccelli [4] and Cohen [46] . . . 132
6.7.2 Preliminary results . . . 135
6.7.3 Stability of discrete models . . . 136
6.7.4 Stability of general models . . . 138
6.8 Converse Theorems . . . 140
6.8.1 Finite models in $\mathbb{Q}_{max}^{k\times k}$ . . . 140
6.8.2 General models . . . 142
6.9 Complementary Results . . . 144
6.9.1 Without irreducibility . . . 144
6.9.2 Boundedness . . . 146
6.10 Appendix . . . 147
6.10.1 Loynes scheme . . . 147
6.10.2 Proof of Theorem 6.7.14 . . . 149
6.10.3 Proof of Theorem 6.8.1 . . . 151
6.10.4 Proof of Theorem 6.8.4 . . . 154
6.10.5 Proof of Theorem 6.8.5 . . . 155

7 Illustration of Multiple Stationary Regimes **159**
7.1 Introduction . . . 160
7.2 Multiple Stationary Regimes . . . 160
7.3 Quantitative Results . . . 164
7.3.1 Systems with two matrices . . . 164
7.3.2 Other systems . . . 169

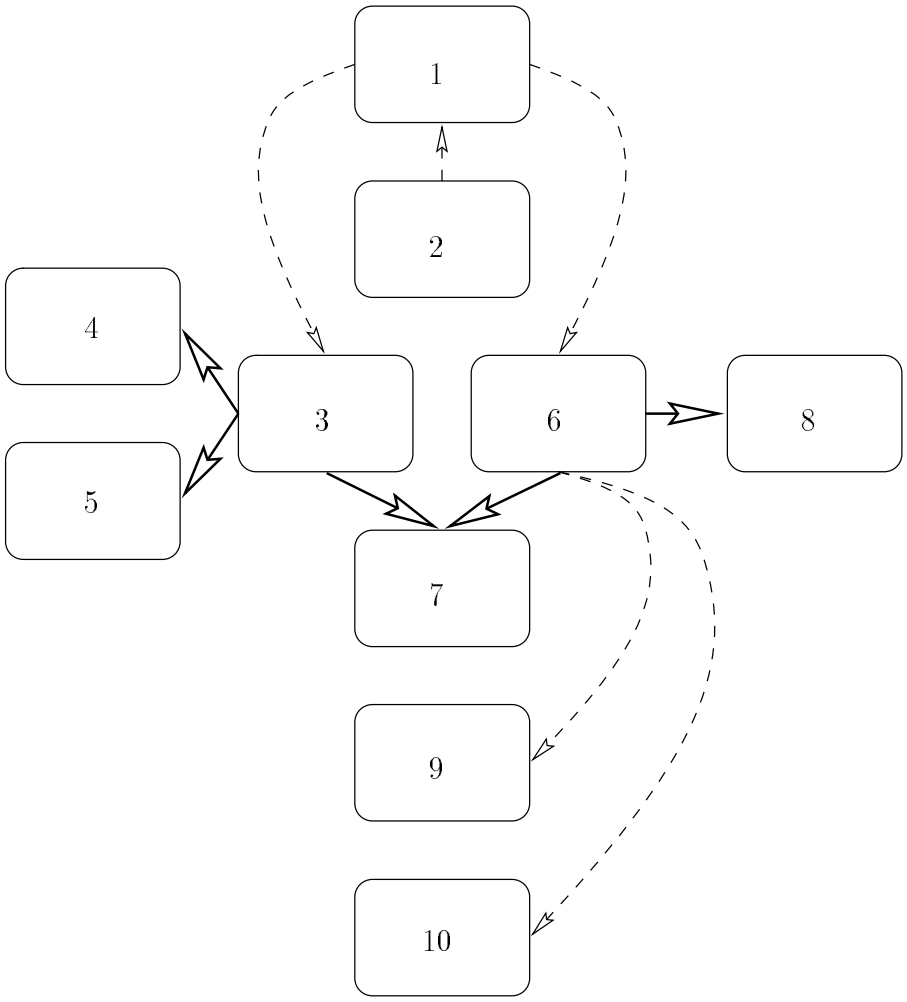
8 Application to Stochastic Event Graphs **170**
8.1 Event Graph . . . 171
8.2 Stochastic Event Graph . . . 172

9 Task Resource Models and (max,+) Automata 176
9.1 Introduction . . . 177
9.2 Basic Task Resource Model . . . 177
9.2.1 General presentation . . . 177
9.2.2 Linear representation over the (max,+) semiring . . . 179
9.2.3 Interpretation in terms of hypergraphs . . . 182
9.2.4 Interpretation in terms of heaps of pieces . . . 182
9.3 Performance Evaluation . . . 183
9.3.1 Stochastic case . . . 183
9.3.2 Optimal case and worst case . . . 186
9.4 Extensions of Task Resource Models . . . 188
9.4.1 Tetris game . . . 188
9.4.2 Gagny's triangle . . . 190
9.5 Appendix . . . 191

III Systèmes Non-Linéaires Stochastiques 194

10 Ergodic Theory of Stochastic Operators and Discrete Event Networks 195
10.1 Introduction . . . 196
10.2 Probabilistic Tools . . . 197
10.3 Application to Stochastic Operators . . . 198
10.3.1 Subadditivity . . . 198
10.3.2 Projective boundedness . . . 200
10.4 Application to Stochastic Discrete Event Networks . . . 202
10.4.1 Discrete event networks . . . 202
10.4.2 The monotone–separable framework . . . 203
10.4.3 Open discrete event networks . . . 205
10.4.4 Stochastic discrete event networks . . . 205
10.5 Relations Between Operators and Networks . . . 206
10.6 Basic Example and Probabilistic Tools . . . 207
10.7 Application to Stochastic Discrete Event Networks . . . 208
10.7.1 Open discrete event networks . . . 208
10.7.2 General discrete event networks . . . 210
10.8 Application to Stochastic Operators . . . 211
10.8.1 Monotonicity . . . 211
10.8.2 Fixed point . . . 214
10.9 Models Entering the Framework . . . 219
10.9.1 Stochastic operators . . . 219
10.9.2 Discrete event networks . . . 222

Plan



Les flèches en trait plein correspondent à un ordre conseillé de lecture. Les flèches en pointillé indiquent des relations entre chapitres qui peuvent néanmoins être lus indépendamment.

Chapitre 0

Introduction

0.1 Systèmes à Evénements Discrets

La motivation première de ce travail est l'étude des *systèmes à événements discrets* (SED). De tels systèmes peuvent se définir de façon négative par opposition aux systèmes classiques dont l'évolution est continue et décrite par des équations différentielles. Dans un SED, les transformations sont déclenchées par des "événements" ponctuels, typiquement l'arrivée d'un client, d'un signal ou l'achèvement d'une tâche. Ces événements donnent lieu à des phénomènes de synchronisation et de concurrence.
De tels systèmes apparaissent de façon naturelle dans la modélisation d'un certain nombre de systèmes physiques. On peut mentionner les exemples suivants.

- Systèmes informatiques, architecture interne des ordinateurs.
- Réseaux de télécommunications, réseaux de transport.
- Systèmes de production : lignes d'assemblage, ateliers flexibles.

Pour décrire ou étudier ces SED, il existe de nombreux modèles et techniques mathématiques. On s'intéressera plus particulièrement aux suivants :

- Réseaux de files d'attente.
- Réseaux de Petri et automates.
- Systèmes dynamiques dans des algèbres non-conventionnelles.

Réseaux de files d'attente, réseaux de Petri et automates seront utilisés comme formalismes de description et de représentation graphique des objets étudiés. L'essentiel de l'analyse aussi bien quantitative que qualitative portera sur les systèmes dynamiques.

0.2 Approche algébrique

La dynamique est codée par des équations de récurrence dans des structures algébriques non-conventionnelles telles l'algèbre (max,+). Une telle approche permet de mettre en œuvre un ensemble de techniques et de résultats proprement algébriques et similaires à ceux utilisés en

algèbre classique. Cette approche permet aussi de s'affranchir de certaines des caractéristiques propres au modèle. Les résultats seront obtenus au niveau algébrique quitte à être ensuite traduits.
Cette démarche est illustrée par les chapitres 6 et 8. Dans le chapitre 6, on démontre des résultats ergodiques pour les produits de matrices aléatoires dans l'algèbre (max,+). Le chapitre 8 est consacré à l'application de ces résultats aux graphes d'événements stochastiques.
La modélisation algébrique permet également de mettre en œuvre des techniques de simulation efficaces. Ce domaine est abordé dans le chapitre 2, §2.7.

Détaillons les différents modèles algébriques considérés.

Systèmes (max,+) linéaires On considère les systèmes décrits par une équation de la forme "$x(n+1) = A \otimes x(n)$", où $x(n)$ et $x(n+1)$ sont des vecteurs et A une matrice carrée. Le produit matrice vecteur $\otimes$ s'interprète en remplaçant les opérations "+" et "$\times$" par "max" et "+" respectivement. De tels systèmes sont les analogues dans une structure algébrique différente des systèmes linéaires classiques.

Opérateurs monotones et homogènes On considère des fonctions (ou opérateurs) $T : \mathbb{R}^k \to \mathbb{R}^k$ vérifiant des propriétés de monotonie, i.e. $x \geqslant y \Rightarrow T(x) \geqslant T(y)$, et d'homogénéité, i.e. $T(x + \lambda\vec{1}) = T(x) + \lambda\vec{1}$, où $\vec{1}$ est le vecteur dont toutes les coordonnées sont égales à 1. De tels opérateurs généralisent strictement les opérateurs (max,+) linéaires ($T(x) = A \otimes x$). Ils correspondent, de façon très imprécise, à des opérateurs (min,max,+,$\times$) linéaires.

Réseaux monotones et séparables De façon schématique, on appelle ici réseau un système où un processus ponctuel (arrivée de clients, de tâches, ...) déclenche un ensemble d'événements internes. Sous des hypothèses adéquates de monotonie et de séparabilité, on obtient une classe de systèmes généralisant strictement les opérateurs monotones et homogènes.

0.3 Systèmes Déterministes-Stochastiques

Les systèmes cités précédemment ont été, historiquement, d'abord étudiés sous des hypothèses déterministes. De nombreux arguments, variabilité intrinsèque des phénomènes étudiés ou impossibilité de les évaluer avec précision, plaident en faveur d'extensions stochastiques de ces mêmes systèmes.

Cette thèse se décompose en deux moitiés. La première (Partie I) est consacrée aux systèmes déterministes et la seconde (Parties II et III) aux systèmes stochastiques.

Dans l'étude déterministe, on s'intéresse à des problèmes d'optimisation et de conception (représentation minimale) ainsi qu'à une analyse fine des comportements asymptotiques (bassins d'attraction des points limites).

Dans l'étude stochastique, on s'intéresse également au comportement limite par l'intermédiaire d'une analyse de la stabilité. Par stabilité d'un SED aléatoire, on entend l'existence d'un régime stationnaire et la convergence vers ce régime pour des quantités telles que le débit, le nombre de tâches (clients, données, ...) ou le temps d'attente dans le système. Plus précisément, on

considère deux notions de convergence, la convergence faible ou étroite (weak convergence) et la convergence en variation (total variation convergence).
Soit $(E, \mathcal{E})$ un espace *polonais* (i.e. métrique, complet et séparable) muni de sa tribu borélienne. On note $\mathcal{M}(E)$ l'ensemble des mesures de probabilité sur $(E, \mathcal{E})$ et $C_b(E)$ l'ensemble des fonctions continues et bornées de E vers $\mathbb{R}$. Soit $\{P_n, n \in \mathbb{N}\}$ et $\{P'_n, n \in \mathbb{N}\} \in \mathcal{M}(E)$.

Convergence	*Définition*	*Théorème de représentation*
étroite	$\forall f \in C_b(E),\ \int f dP_n - \int f dP'_n \to 0$	$X_n \to X'_n,\ P-p.s.$
variation	$\sup_{A\in\mathcal{E}} \left(\int 1_A dP_n - \int 1_A dP'_n \right) \to 0$	$\forall n \geqslant N,\ X_n = X'_n$ et $P(N < \infty) = 1$

Les théorèmes de représentation doivent s'interpréter comme l'existence d'un espace de probabilité $(\Omega, \mathcal{F}, P)$ et de variables aléatoires $X_n, X'_n, n \in \mathbb{N}$ définies sur cet espace et de lois de probabilités respectives $P_n, P'_n, n \in \mathbb{N}$. Pour la convergence en variation, ce résultat est du à Goldstein (voir Lindvall [102]). Ils permettent une approche trajectorielle et intuitive. La convergence faible correspond au couplage asymptotique et la convergence en variation au couplage en temps fini.

Mentionnons quelques-unes des raisons justifiant l'intérêt d'une étude de stabilité.

- En général, on attend d'un réseau que le nombre de tâches en attente ou la durée d'exécution d'une tâche restent finis au cours du fonctionnement. La zone de stabilité du réseau correspond à sa zone de bon fonctionnement. Lors de la conception d'un réseau, un des critères communément utilisés est la maximisation de la zone de stabilité.

- L'étude de la stabilité a connu un vif regain d'intérêt avec la découverte récente de réseaux non-stables sous les conditions usuelles de charge ("$\rho < 1$"), voir Rybko et Stolyar [125] ou Dumas [59] (zone de stabilité non-convexe et possédant une frontière quadratique). A l'inverse, et de façon complémentaire, notre approche va consister à définir des classes de systèmes stables sous des conditions naturelles.

- La connaissance de la dépendance du régime stationnaire en fonction de la condition initiale (analyse fine de stabilité) est utile pour le contrôle ou la simulation d'un réseau. On choisira une condition initiale en fonction du mode de fonctionnement que l'on cherche à obtenir ou simuler.

Dans le cadre de l'analyse des systèmes stochastiques, on supposera parfois que les suites de variables aléatoires sont i.i.d. (indépendantes et identiquement distribuées). Le plus souvent, cependant, on mènera l'étude sous des hypothèses de type stationnaire et ergodique. Les justifications en sont multiples.

- D'un point de vue théorique, il est toujours satisfaisant de démontrer les résultats sous des hypothèses minimales.

- Le cadre stationnaire et ergodique permet de prendre en compte les phénomènes de *périodicité*, telle la dépendance des variables aléatoires en fonction du moment de la journée ou de l'année.

- Des études expérimentales menées récemment sur des données réelles (trafic sur réseau local Ethernet) ont permis de mettre en lumière des phénomènes de longues dépendances et même d'auto-similarité (nature fractale) du trafic, voir [141]. De tels trafics ne sont évidemment pas i.i.d. Il apparaît par contre qu'ils peuvent être efficacement représentés à l'aide de processus stationnaires et ergodiques tels les mouvements browniens fractionnaires.

0.4 Résumé et Contributions

Le chapitre 0 consiste en une introduction générale. Il contient un résumé de la thèse.

Partie I

Le premier chapitre, le seul écrit en français[1], propose une présentation d'un ensemble de systèmes représentables sous forme (max,+) linéaire. Ce chapitre a aussi pour fonction de montrer que les méthodes et résultats présentés par la suite peuvent s'appliquer à des domaines autres que les SED, tels l'économie mathématique ou la mécanique statistique.

Le chapitre 2 est consacré à l'étude de systèmes d'Equations Récurrentes Uniformes. Un tel système peut être vu comme un graphe infini et périodique de calculs à effectuer. L'objectif est de minimiser le nombre de cases mémoires nécessaires pour mener à bien ce calcul. Le modèle étudié est très général et se situe en amont des modèles considérés par la suite.

Le chapitre 3 traite le cas des systèmes (max,+) linéaires déterministes de dimension 3. On propose un nouvel outil de description du comportement spectral. Il s'agit de la représentation graphique des vecteurs propres et des domaines d'attraction dans un "espace projectif additif".

Le chapitre 4 complète le chapitre précédent. Il s'agit d'illustrer à l'aide de la représentation graphique des problèmes propres à l'ordonnancement cyclique dans les systèmes de production.

Le chapitre 5 propose un algorithme permettant de calculer tous les vecteurs propres *et* tous les régimes périodiques d'une matrice (max,+).

Partie II

Le chapitre 6 étudie les produits de matrices aléatoires dans l'algèbre (max,+). On obtient des conditions nécessaires et suffisantes pour que de tels systèmes couplent en temps fini avec un unique régime stationnaire.

Le chapitre 7 vient en complément du chapitre précédent. Il illustre les phénomènes de régimes stationnaires multiples à l'aide de l'outil graphique du chapitre 3.

Le chapitre 8 applique les résultats du chapitre 6 au problème de la stabilité des graphes d'événements stochastiques.

Le chapitre 9 introduit un modèle de ressources partagées, dit modèle Tâche Ressource. Celui-ci se représente sous la forme d'automate (max,+), c'est-à-dire de système (max,+) linéaire dont la dynamique est indexée par des lettres. On étudie sa stabilité à l'aide des résultats du chapitre 6.

[1] Editor's note : almost the only one, chapter 0 is also in french.

Partie III

Le chapitre 10 propose un cadre général pour aborder le problème de la stabilité des systèmes à événements discrets stochastiques. D'une part, on étudie les itérés d'opérateurs aléatoires homogènes et monotones, d'autre part les réseaux à événements discrets monotones et séparables. On propose des théorèmes ergodiques dits du premier et du second ordre. Ces résultats généralisent ceux obtenus au chapitre 6. On traite ensuite plus spécifiquement le modèle des réseaux de Jackson.

Contributions

	Fermé (autonome)	Ouvert (non-autonome)
Systèmes (max,+) linéaires déterministes	1 ordre : Cuninghame-Green [49] 2 ordre : Cohen, Moller, Quadrat, Viot [45] Appr. graph. J.M. Ch. 3	1 ordre : Cohen, Dubois, Quadrat, Viot [43] 2 ordre : Wende & al [139] Gaubert [67]
Systèmes (max,+) linéaires aléatoires	1 ordre : Cohen [46] 2 ordre : J.M. Ch. 6	1 ordre : Baccelli [4] 2 ordre : Baccelli [4]
Opérateurs monotones homogènes aléatoires	1 ordre : Vincent [136] 2 ordre : Baccelli, J.M. Ch. 10	1 ordre : Vincent [136] 2 ordre : Borovkov [22] Brandt, Franken, Lisek [32]
Systèmes monotones séparables aléatoires	1 ordre : Baccelli, Foss, J.M. Ch. 10 2 ordre : $\emptyset$	1 ordre : Baccelli, Foss [10] 2 ordre : Baccelli, Foss [10]

Figure 1: Principales contributions.

On propose, en figure 1, un schéma simplifié permettant de situer une partie des contributions de la thèse. Seuls certains des chapitres sont mentionnés, les autres s'y rattachent (voir le plan précédant ce chapitre) ou portent sur des domaines connexes (modélisation : Ch. 1,9, et optimisation : Ch. 2).
Les résultats originaux de cette thèse portent principalement sur les systèmes fermés. Considérons un réseau avec des serveurs et des clients. Ce réseau sera dit ouvert s'il existe un flux de clients arrivant, circulant puis quittant le réseau. Au contraire, il sera dit fermé si le nombre de clients dans le réseau est fixe. Ces définitions demanderont bien sûr à être précisées.
En langage algébrique, les résultats du *premier ordre* sont ceux relatifs aux valeurs propres et les résultats du *second ordre* aux vecteurs propres. En langage réseau, les résultats du premier ordre portent sur le débit et les résultats du second ordre sur le nombre de tâches ou le temps d'attente dans le système. Les références données en figure 1 sont très incomplètes. Pour plus de détails, on se reportera aux chapitres correspondants.

0.5 Problèmes Non Résolus

Les problèmes ouverts évoqués ici constituent autant de pistes de recherches dans la continuité de ce travail.

- Pouvoir de modélisation des systèmes étudiés.

Le pouvoir d'expression des systèmes (max,+) linéaires est maintenant bien compris, voir chapitre 1. En terme de réseaux de Petri, ils correspondent à la sous-classe des graphes d'événements. Il n'en est pas du tout de même pour la classe des systèmes (min,max,+) linéaire ou celle des réseaux monotones-séparables (voir chapitre 10). En particulier, l'intersection entre ces systèmes et la classe des réseaux de Petri reste à éclaircir.

- Etude des systèmes (max,+) linéaires stochastiques de dimension infinie.

Une théorie spectrale relativement complète existe pour les systèmes (max,+) linéaires déterministes de dimension infinie, voir [107]. Il n'existe par contre pas, à notre connaissance, de théorie ergodique pour les versions stochastiques de ces systèmes. Une application potentiellement intéressante serait l'étude du modèle d'exclusion asymétrique, exemple de système de particules en interaction [101]. Ce modèle apparaît entre autre dans le fameux problème de la suite infinie de files en tandem et peut se modéliser sous la forme d'un système (max,+) linéaire de dimension infinie dénombrable, voir [5].

- Etude des opérateurs monotones-homogènes de dimension infinie.

Les réseaux de Jackson sont un exemple de réseaux monotones-séparables pouvant s'interpréter comme un opérateur (min,max,+) à dépendances non-bornées (Lemme 10.9.5). De façon générale, on peut se demander s'il y a équivalence entre la classe des réseaux monotones-séparables et celle des opérateurs monotones-homogènes. La réponse à une telle question constituerait un premier pas dans l'étude des opérateurs de dimension infinie.

Partie I

Systèmes Linéaires Déterministes

Chapitre 1

Exemples de Systèmes (max,+) Linéaires

Une importante caractéristique des systèmes (max,+) linéaires est leur simplicité mathématique. Il s'agit en effet de considérer des produits de matrices mais dans une structure algébrique non usuelle. Avoir un modèle théorique simple est certainement un atout, encore faut-il que cela s'accompagne d'une puissance de modélisation raisonnable. Les systèmes linéaires dans l'algèbre (max,+) réalisent ce compromis. Ce chapitre illustre le second aspect, le pouvoir de modélisation.

Ce chapitre doit paraître dans la *Revue Scientifique et Technique de la Défense*. Mes remerciements vont à Camille Terray, Sophie Lefebvre-Barbaroux et Alain Jean-Marie pour leur relecture attentive d'une première version.

1.1 Introduction

On présente un ensemble de modèles dont la caractéristique commune est de pouvoir être représentés par un système linéaire dans l'algèbre (max,+).
On ne cherchera pas à établir un catalogue exhaustif de modèles mais plutôt à illustrer la variété des domaines où ils apparaissent.
Ce chapitre a également pour objet, à un modeste niveau, d'éviter l'écueil consistant à se limiter à une vision ou à un domaine d'application. La multiplicité des domaines est susceptible de suggérer et d'éclairer des problèmes différents et complémentaires.

Le chapitre est organisé de la façon suivante. En section 1.2, on présente le modèle mathématique théorique. Les quatre sections suivantes sont indépendantes les unes des autres. Elles sont consacrées à quatre types d'applications : les graphes d'événements (sous-classe de réseaux de Petri), les automates temporisés, l'optimisation dynamique et enfin le modèle de Frenkel-Kontorova en mécanique statistique. Ces différents modèles ne sont pas originaux en ce sens qu'ils étaient déjà répertoriés, au moins par les membres de leur communauté scientifique respective, en tant que systèmes (max,+) linéaires. La seule exception est le modèle de réseau de Petri avec choix présenté en section 1.4.2 et qui est inspiré d'un travail en préparation [71].

1.2 Modèle Mathématique

On considère le système d'équations récurrentes :

$$\begin{aligned} x_i(n+1) &= \max_{1\leqslant j\leqslant k}(A_{ij}+x_j(n)),\ i=1,\dots,k\ , \\ x_i(0) &= (x_0)_i\ . \end{aligned} \tag{1.1}$$

Les quantités $x_i(n)$ et A_{ij} appartiennent à $\mathbb{R}\cup\{-\infty\}$. On veut étudier la suite de vecteurs $\{x_1(n),\dots,x_k(n)\}$.

Il est fructueux de récrire l'équation (1.2) en utilisant une notation matrice-vecteur.

Définition 1.2.1. *L'algèbre (max,+) ou $\mathbb{R}_{\max}$ est l'ensemble $\mathbb{R}\cup\{-\infty\}$, muni de la loi* max, *notée additivement (i.e. $a\oplus b=\max(a,b)$) et de la loi +, notée multiplicativement (i.e. $a\otimes b=a+b$).*

On définit la matrice A de dimension $k\times k$ dont les coordonnées sont A_{ij}. On définit également le vecteur colonne $x(n)=(x_1(n),\dots,x_k(n))'$ ainsi que le vecteur de conditions initiales x_0. Avec ces nouvelles notations, l'équation (1.2) prend une forme simple et agréable :

$$x(n+1) = A\otimes x(n),\ \text{ i.e. }\ x(n+1)=A^{\otimes n+1}\otimes x_0\ . \tag{1.2}$$

Le produit matriciel est défini de façon naturelle en remplaçant simplement les opérations + et $\times$ de l'algèbre usuelle par $\oplus$ et $\otimes$. Soit A et B deux matrices de taille appropriée, on a $(A\oplus B)_{ij}=A_{ij}\oplus B_{ij}=\max(A_{ij},B_{ij})$, $(A\otimes B)_{ij}=\bigoplus_k A_{ik}\otimes B_{kj}=\max_k(A_{ik}+B_{kj})$. Dans la suite, on omettra souvent le symbole $\otimes$, remplaçant par exemple $A\otimes B$ par AB.

Une généralisation naturelle, et souvent essentielle dans la modélisation de systèmes physiques, consiste à considérer que la matrice A n'est pas constante. L'équation (1.2) prend dès lors la forme suivante :

$$x(n+1) = A(n)x(n), \text{ i.e. } x(n+1) = A(n)\cdots A(0)x_0\,. \tag{1.3}$$

La suite $\{A(n)\}$ est donnée de façon exogène. En général, ce sera le cas dans les chapitres de la partie II, on suppose que la suite $\{A(n)\}$ est une suite i.i.d. ou stationnaire ergodique de matrices aléatoires.

Pour les systèmes d'équations (1.2) ou (1.3), on définit deux types de limites asymptotiques :

• Les limites dites du premier ordre : $\lim_n ||x(n)||/n$, $\lim_n x_i(n)/n$ (division dans l'algèbre usuelle).

• Les limites dites du second ordre : $\lim_n\ x_i(n+1) - x_i(n)$, $\lim_n\ x_j(n) - x_i(n)$, $\forall i, j$.

Dans le cas d'un système déterministe, i.e. $A(n) \equiv A$, ces limites sont directement reliées aux éléments propres de la matrice A. Ceux-ci sont définis de la façon suivante. On cherche $\lambda \in \mathbb{R}$ et $x \in \mathbb{R}^k$ solutions de l'équation spectrale :

$$\begin{aligned} \max_{1\leqslant j\leqslant k}(A_{ij} + x_j) &= \lambda + x_i, \quad i = 1, \ldots, k \\ A \otimes x &= \lambda \otimes x\,, \end{aligned} \tag{1.4}$$

Par analogie avec l'algèbre classique, on appelle λ une valeur propre et x un vecteur propre de la matrice A. Une particularité importante de l'algèbre (max,+) est l'existence d'une **unique** valeur propre pour une matrice A irréductible [1]. Par contre, il peut y avoir une multiplicité de vecteurs propres ainsi que des régimes périodiques[2]. Pour plus de détails, voir chapitre 3.
On verra dans la suite quelle interprétation donner aux limites du premier et du deuxième ordre en fonction du modèle physique considéré.

Algèbre de chemin L'algèbre (max,+) est souvent appelée une algèbre de chemin pour la raison suivante. A toute matrice $A \in \mathbb{R}_{max}^{k\times k}$, on peut associer un graphe à k nœuds comportant un arc de j vers i de poids A_{ij} si $A_{ij} > -\infty$. Le terme $A_{ij}^{\otimes p}, p \in \mathbb{N}$, s'interprète alors comme le poids maximum des chemins de longueur (mesurée en nombre de nœuds) p joignant j à i. Cette interprétation sera utile par la suite.

1.3 Réseaux de Petri

L'évaluation de performances de systèmes informatiques et de télécommunications nécessite une modélisation préalable de ces modèles. Cette modélisation utilise souvent le paradigme des Systèmes Dynamiques à Événements Discrets (SED). Pour de plus amples détails sur le champ d'application des SED, on pourra se référer au numéro spécial de *Proceedings of the IEEE* [88]

[1] i.e. $\forall i, j, \exists n$ s.t. $A_{ij}^n > -\infty$.

[2] Un régime périodique est un ensemble fini $x^1, \ldots, x^d$ tel que $Ax^1 = \lambda x^2, Ax^2 = \lambda x^3, \ldots, Ax_d = \lambda x^1$.

ou aux ouvrages récents de Baccelli, Cohen, Olsder & Quadrat [8] et de Glasserman & Yao [75]. Une excellente référence récente en français est Gaubert [68].
Parmi les formalismes de modélisation des SED les plus largement utilisés, on peut mentionner les GSMP (Generalized Semi-Markov Process) et les réseaux de Petri. Il est montré dans [75] qu'un GSMP vérifiant certaines propriétés de convexité et d'homogénéité peut être représenté sous la forme d'une équation de type (1.3).
On va insister dans la suite de cette section sur le second formalisme, celui des réseaux de Petri. Ils ont été introduit par Carl Petri en 1962 dans sa thèse soutenue à l'université de Bonn [117]. Notre but n'est pas de présenter le formalisme dans toute sa richesse mais d'arriver le plus vite possible à la sous-classe qui nous intéresse, celle des graphes d'événements temporisés. Pour une description plus complète, on se reportera à Murata [112] ou Brams [30].

1.3.1 Présentation générale

Un réseau de Petri est défini comme le quadruplet $\mathcal{G} = (\mathcal{P}, \mathcal{T}, \mathcal{F}, \mathcal{M})$, où :

- $\mathcal{P}$ est un ensemble fini. Ses éléments sont appelés places.
- $\mathcal{T}$ est un ensemble fini. Ses éléments sont appelés transitions.
- $\mathcal{F} \subseteq (\mathcal{P} \times \mathcal{T}) \cup (\mathcal{T} \times \mathcal{P})$ définit une relation de dépendance entre places et transitions.
- $\mathcal{M}$ est une fonction de $\mathcal{P}$ dans $\mathbb{N}$. L'entier $\mathcal{M}(p)$ est appelé le marquage de la place $p \in \mathcal{P}$.

Un réseau de Petri peut être interprété comme un graphe orienté. Les nœuds sont de deux types : les places et les transitions. Un élément de $\mathcal{F}$ est un arc reliant une place et une transition ou bien une transition et une place. Il devient dès lors naturel de parler de "places d'entrée" (d'une transition), de "transitions de sortie" (d'une place), etc.
Un formalisme graphique spécifique est associé aux réseaux de Petri. Les places sont représentées par des cercles et les transitions par des barres. Un marquage $n = \mathcal{M}(p)$ est représenté par n jetons dans la place p. Un exemple de ce formalisme est fourni par la figure 1.1.

Un réseau de Petri est également un objet dynamique. Le triplet $(\mathcal{P}, \mathcal{T}, \mathcal{F})$ n'est jamais modifié mais le marquage évolue suivant une règle intitulée *tir d'une transition*. Cette règle est définie comme suit :

1. Une transition t est dite *habilitée* si il y a au moins un jeton dans chaque place d'entrée de t.
2. Une transition t peut *tirer* si et seulement si elle est habilitée.
3. Le *tir* de t enlève un jeton dans chaque place d'entrée et ajoute un jeton dans chaque place de sortie.

La règle de tir est illustrée par la figure 1.1. On a représenté un réseau de Petri avant et après le tir de la transition t. Il est possible qu'un même jeton puisse participer au tir de plusieurs transitions (il faut que la place correspondante ait plusieurs transitions en sortie). Dans ce cas, on dit être en présence d'un *choix*. C'est le cas pour la figure 1.1, la transition t' étant également initialement habilitée.

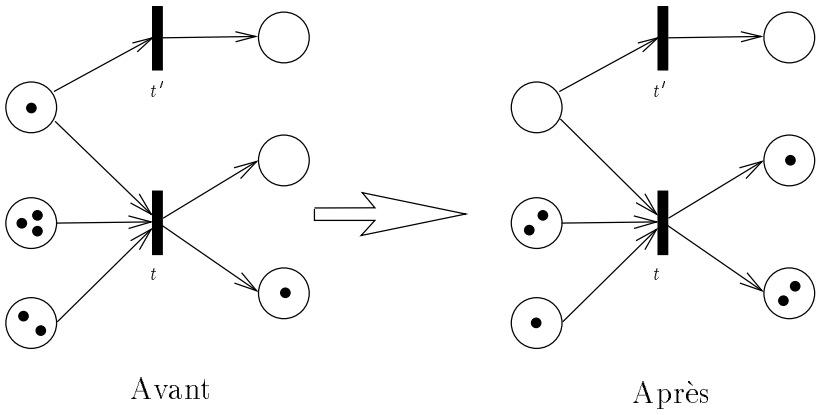


Figure 1.1: Réseau de Petri. Avant : les transitions t et t' sont habilitées. Après : la transition t vient d'être tirée. Aucune transition n'est habilitée.

Le plus souvent, les places représentent des conditions (présence de clients, de ressources, ...) et les transitions des événements.

Temporisation Un réseau de Petri est à l'origine un objet logique. Cependant, la temporisation de cet objet s'est avérée une approche très riche permettant en particulier l'évaluation de performance du système modélisé.
A chaque transition t (resp. place p), on associe $\sigma_t \in \mathbb{R}_+$ (resp. $\sigma_p \in \mathbb{R}_+$).

- Le réel σ_p correspond au *temps de séjour* d'un jeton en place p. Plus précisément, si un jeton arrive en place p à l'instant u, il ne sera disponible pour l'habilitation des transitions (i.e. l'étape 1 de la règle de tir) en sortie de t qu'à l'instant $u + \sigma_p$.
- Le réel σ_t correspond au *temps de tir* de la transition t. Il s'agit du temps qui s'écoule entre le début et la fin du tir de t. Pendant la durée σ_t, les jetons participant au tir sont "gelés" et ne peuvent participer à une autre habilitation ou à un autre tir.

Dans la suite de cette section, on va se restreindre à une sous-classe de réseaux de Petri, les *graphes d'événements* (parfois appelées *graphes marqués* ou *réseaux de Petri sans décisions*).

Définition 1.3.1 (Graphe d'événements). *Un graphe d'événements est un réseau de Petri dans lequel chaque place a exactement une transition en entrée et une en sortie.*

Les graphes d'événements permettent de modéliser la synchronisation mais excluent les choix. On reviendra sur les réseaux de Petri avec choix en section §1.4.2.

Un graphe d'événements est *vivant* si chaque circuit contient au moins un jeton. Lorsque cette propriété n'est pas vérifiée, le réseau se bloque après un nombre fini de tirs. Dans toute la suite, il est implicite que l'on ne considère que des graphes d'événements vivants.

Considérons un graphe d'événements temporisé $\mathcal{G} = (\mathcal{P}, \mathcal{T}, \mathcal{F}, \mathcal{M}, \{\sigma_t, t \in \mathcal{T}\}, \{\sigma_p, p \in \mathcal{P}\})$. On cherche à le représenter sous forme d'un système (max,+) linéaire. Pour ce faire, commençons

par réaliser une transformation du graphe. Chaque place p comprenant $M(p) > 1$ jetons est remplacée par $M(p)$ places en série avec exactement 1 jeton par place. Les nouvelles transitions et les nouvelles places ont des temps de tir et de séjour identiquement égaux à 0. Cette transformation a été illustrée en figure (1.2).

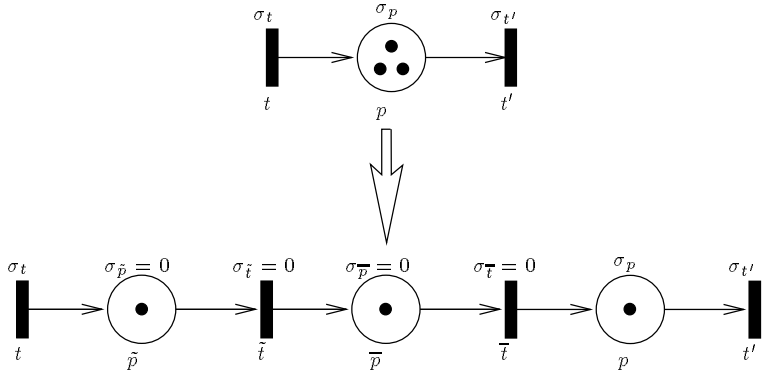


Figure 1.2: Eclatement d'une place comprenant plusieurs jetons.

Pour simplifier, on conserve les notations $\mathcal{G} = (\mathcal{P}, \mathcal{T}, \mathcal{F}, \mathcal{M}, \{\sigma_t\}, \{\sigma_p\})$ pour le nouveau graphe. On définit une suite de vecteurs $\{x(n), n \in \mathbb{N}\}$ de dimension $k = |\mathcal{T}|$, le nombre de transitions. Le terme $x(n)_i$ correspond à la date de début du n-ième tir de la transition t_i. Pour des raisons de convenance, on note ${}^\bullet i \subset \mathcal{T}$ l'ensemble des transitions en entrée de la transition t_i, i.e. l'ensemble des t_j tels que $\exists p \in \mathcal{P} \mid (t_j, p) \in \mathcal{F}, (p, t_i) \in \mathcal{F}$. D'autre part, si il existe une place entre les transitions t_i et t_j, on la note p_{ij}.
Soit $j \in {}^\bullet i$. On considère le n-ième jeton produit par la transition t_j. En raison de la propriété $\mathcal{M}(p) \leqslant 1, \forall p$, deux cas seulement sont possibles : soit ce jeton habilite le n-ième tir de t_i (si $\mathcal{M}(p_{ji}) = 0$), soit il habilite le $(n+1)$-ième tir de t_i (si $\mathcal{M}(p_{ji}) = 1$). La règle de tir au niveau de la transition t_i se traduit par l'équation suivante :

$$x_i(n+1) \geqslant \max_{j \in {}^\bullet i} \left\{ x_j(n+1-\mathcal{M}(p_{ji})) + \sigma_{t_j} + \sigma_{p_{ji}} \right\} . \tag{1.5}$$

En particulier l'instant $u = x_j(n+1-\mathcal{M}(p_{ji})) + \sigma_{t_j}$ correspond à la fin du $(\, n+1-\mathcal{M}(p_{ji})\,)$-ième tir de la transition t_j. Le jeton correspondant habilite donc la transition t_i à partir de l'instant $u + \sigma_{p_{ji}}$.

On définit les matrices A_u, $u = 0$ ou 1, de la façon suivante :

$$(A_u)_{ij} = \begin{cases} \sigma_{t_j} + \sigma_{p_{ji}} & \text{si } j \in {}^\bullet i \text{ et } \mathcal{M}(p_{ji}) = u \\ -\infty & \text{sinon} \end{cases} .$$

L'équation (1.5) se récrit sous la forme :

$$x(n+1) \geqslant A_0 x(n+1) \oplus A_1 x(n) . \tag{1.6}$$

Introduisons la matrice $A_0^\star = \bigoplus_{p=0}^{\infty} A_0^p = \bigoplus_{p=0}^{k} A_0^p$, où A_0^0 est la matrice identité I définie par $I_{ii} = 0$ et $I_{ij} = -\infty, i \neq j$. On montre aisément que cette matrice est l'inverse formel de $I - A_0$, i.e. $A(0)^\star(I - A(0)) = (I - A(0))A(0)^\star = I$. On peut montrer

$$x(n+1) \geqslant A_0^* A_1 x(n) \; . \tag{1.7}$$

Le passage de (1.6) à (1.7) illustre la puissance formelle de la représentation algébrique sous forme matricielle. En parcourant le chemin inverse de celui réalisé précédemment, on peut associer à l'équation (1.7) un graphe d'événements avec **exactement** un jeton par place. Il est instructif de comparer ce nouveau graphe avec le graphe initial. Un exemple simple est proposé en figure 1.3.

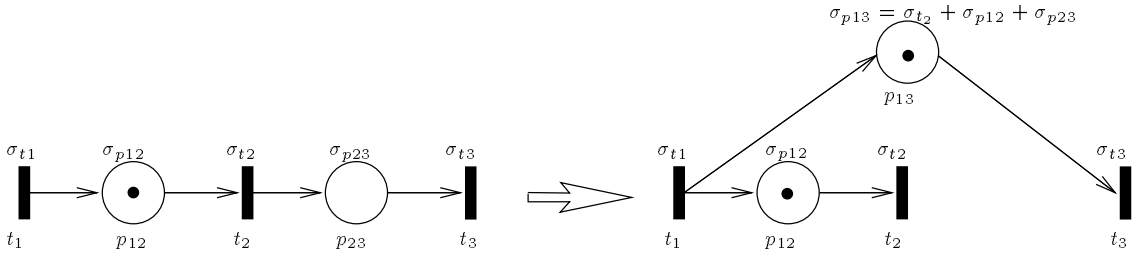


Figure 1.3: La multiplication par A_0^* a pour effet de court-circuiter les chemins sans jetons.

Evolution au plus tôt On est en général particulièrement intéressé par l'évolution au plus tôt du graphe d'événements, c'est-à-dire qu'une transition tire dès qu'elle est habilitée. Cette évolution correspond à la solution minimale de (1.7), c'est-à-dire à l'équation (max,+) linéaire :

$$x(n+1) = A_0^* A_1 x(n) \; . \tag{1.8}$$

Une propriété remarquable des graphes d'événements est que le marquage est préservé lorsque l'on tire chaque transition une fois, voir chapitre 8 proposition 8.1.4. Cela explique que la récurrence (1.8) soit uniforme en n, le marquage à l'"instant" $x(n)$ (après n tirs de chaque transition) étant exactement le marquage initial $\mathcal{M}$. Il est toutefois important de comprendre qu'il s'agit d'un marquage "virtuel", au sens où il apparaît à un "instant" $x(n)$ correspondant à des dates différentes des horloges associées à chaque transition. En général, une photographie du réseau à un instant $t > 0$ ne permettra pas d'observer le marquage $\mathcal{M}$. En effet il faudrait pour cela qu'il existe une occurrence n et un réel t tel que $(x_1(n), \ldots, x_k(n)) = (t, \ldots, t)$, ce qui n'est pas toujours le cas.

Remarque 1.3.2. Il existe des systèmes (max,+) linéaires autres que celui de l'équation (1.7) et qui représentent l'évolution du graphe d'événements. En particulier, on aurait pu considérer des dates de fin de tir au lieu de dates de début de tir. De façon plus essentielle, il existe des méthodes plus astucieuses d'éclatement des places marquées que celle présentée en figure 1.2. Cela permet d'obtenir une représentation à l'aide d'une matrice de taille inférieure. La recherche d'une représentation de taille minimale constitue l'objet du chapitre 2.

Temporisations aléatoires Une généralisation naturelle est de considérer que les temporisations des places et transitions du graphe d'événements sont données par des suites $\{\sigma_t(n), n \in \mathbb{N}\}$, $t \in \mathcal{T}$, et $\{\sigma_p(n), n \in \mathbb{N}\}$, $p \in \mathcal{P}$, de variables aléatoires. On aimerait pouvoir écrire les analogues des équations (1.5) (1.6) (1.7) et (1.8).
Cependant, dans un graphe d'événements stochastique, il devient possible pour les jetons de se dépasser (il suffit d'imaginer que $\sigma_p(n) \gg \sigma_p(n+1)$). Si on étudie de près l'équation (1.5), on se rend compte qu'elle n'est plus vraie dès lors que des dépassements sont possibles. Pour contourner cette difficulté, il faut se limiter à une sous-classe de graphes d'événements stochastiques, dits PAPS (Premier Arrivé Premier Servi), pour lesquels les dépassements sont impossibles. Sous les hypothèses qui suivent, un graphe d'événements est PAPS.

- Toute transition t_i vérifie une des deux propriétés suivantes : 1) il existe un rebouclage avec un unique jeton, i.e. $i \in {}^\bullet i$ et $\mathcal{M}(p_{ii}) = 1$. 2) le temps de tir est constant, i.e. $\sigma_t(n) = \sigma_t$, $\forall n$.
- Toute place p a un temps de séjour constant, i.e. $\sigma_p(n) = \sigma_p$, $\forall n$.

Il existe d'autres types de conditions permettant d'assurer une évolution de type PAPS. Ainsi un graphe d'événements avec au plus un jeton par circuit sera toujours PAPS, voir l'exemple de la figure 1.5.

Pour un graphe d'événements stochastique PAPS, on définit les suites de matrices $\{A_u(n), n \in \mathbb{N}\}, u = 0$ ou 1 :

$$A_u(n)_{ij} = \begin{cases} \sigma_{t_j}(n+1-\mathcal{M}(p_{ij})) + \sigma_{p_{ji}} & \text{si } j \in {}^\bullet i \text{ et } \mathcal{M}(p_{ji}) = u \\ -\infty & \text{sinon .} \end{cases}$$

L'évolution au plus tôt du graphe d'événements est donné par le système (max,+) linéaire :

$$x(n+1) = A_0(n)^* A_1(n) x(n) . \tag{1.9}$$

On propose, sections §1.3.2 et §1.3.3, deux exemples de systèmes se modélisant sous forme de graphe d'événements et donc de système (max,+) linéaire.

1.3.2 Atelier flexible

Un atelier flexible (job-shop en anglais) est un type particulier de système de production. On dispose d'un nombre fini de machines et d'un nombre fini de types de produit. Chaque produit doit subir une suite d'opérations élémentaires sur différentes machines et ce dans un ordre bien déterminé. Chaque machine travaille de façon séquentielle sur un produit à la fois. Nous avons représenté en figure 1.4 un exemple avec deux machines (M_1, M_2) et deux types de produits (j_1, j_2) représentés par leur *routage* le long des machines.
On suppose qu'il y a un stock infini de chaque type de produit. Dès qu'un produit de type p achève son cycle de production, un nouveau produit p commence le sien.

Pour que le modèle soit complètement spécifié, il faut définir l'ordre de passage, ou *ordonnancement*, des produits sur les machines. On considère dans la suite que cet ordonnancement est

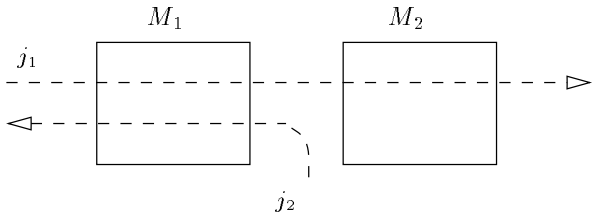


Figure 1.4: Atelier flexible constitué de deux machines et deux types de produits.

fixé (imposé par les contraintes technologiques) et périodique. Suivant les notations de Hillion et Proth [87], on note $\sigma(M)$ l'ordonnancement des produits sur la machine M. Le routage du produit j est noté $\sigma(j)$. Pour le modèle de la figure 1.4 par exemple, on a $\sigma(j_1) = (M_1, M_2)$, $\sigma(j_2) = M_1$. Un ordonnancement possible est $\sigma(M_1) = (j_1 j_1 j_2)(j_1 j_1 j_2) \cdots$ et $\sigma(M_2) = j_1 j_1 \cdots$.

On peut définir la période minimale commune des suites $\sigma(M_i)$. Cette période minimale peut comprendre plusieurs produits de chaque type. Ainsi dans l'exemple cité ci-dessus, la période est 3 et correspond à deux produits j_1 et un produit j_2. L'ensemble des produits correspondant à une période minimale est appelé MPS (Minimal Part Set, Lee [99]).

Un atelier flexible peut être représenté sous forme d'un graphe d'événements, voir [87]. Plutôt que de présenter la construction générale, on se propose de l'illustrer à l'aide d'un exemple.

On considère l'atelier de la figure 1.4 avec l'ordonnancement suivant :

$$\sigma(M_1) = (j_2, j_1)(j_2, j_1) \cdots \quad \text{et} \quad \sigma(M_2) = (j_1) \cdots$$

Le MPS est formé par un produit de chaque type. La représentation sous forme de graphe d'événements de ce système est donnée en figure 1.5.

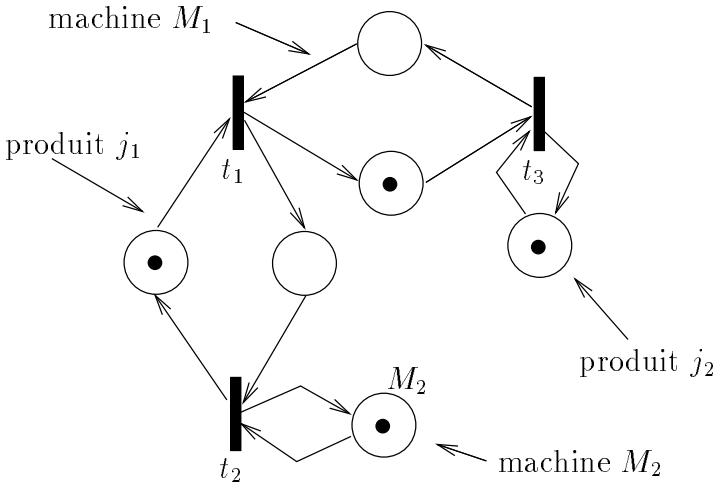


Figure 1.5: Atelier flexible. Modélisation sous forme de graphe d'événements.

La synthèse de ce graphe est réalisée en deux étapes. A chaque produit p du MPS, on associe un cycle comportant un nombre de transitions égal au cardinal de $\sigma(p)$. Une transition correspond au passage du produit p sur une machine. Chaque cycle-produit comporte exactement 1 jeton. Ceci correspond au mécanisme décrit plus haut : dès que la fabrication d'un produit est achevée, la fabrication d'un nouveau produit du même type commence.
Il faut ensuite modéliser le fait que certaines de ces transitions correspondent à la même machine physique. Les transitions correspondant à la machine M_i sont reliées par un cycle suivant l'ordre donné par $\sigma(M_i)$. Chacun de ces cycles-machines comprend exactement un jeton, ceci correspondant au fonctionnement séquentiel des machines.

La méthode générale permettant de décrire l'exécution au plus tôt d'un graphe d'événements par une matrice (max,+) s'applique. Les dateurs $x_1(n), x_2(n)$ et $x_3(n)$ sont associés aux transitions t_1, t_2 et t_3 de la figure 1.5. Soit (α, β, γ) les temps de tir de ces transitions (durée de passage du produit sur la machine). On suppose que les temps de transport entre machines sont nuls. Tous calculs effectués, on obtient le système linéaire suivant :

$$x(n+1) = \begin{pmatrix} \alpha+\gamma & \beta & 2\times\gamma \\ 2\times\alpha+\gamma & \alpha+\beta & \alpha+2\times\gamma \\ \alpha & -\infty & \gamma \end{pmatrix} \otimes x(n) \ .$$

L'inverse de la valeur propre de la matrice, $1/\lambda$, correspond au taux de production de l'atelier. On obtient ici $1/\lambda = (\ \alpha + \max(\beta,\gamma)\ )^{-1}$. Si on contrôle l'atelier en imposant une condition initiale $x(0)$ qui soit un vecteur propre, on obtient un fonctionnement régulier de l'atelier, où les opérations se répètent à l'identique toutes les λ unités de temps.

1.3.3 Files d'attente avec blocage

La notation de Kendall est une façon simple et synthétique de décrire une file d'attente. Avec cette notation, la file la plus simple s'écrit[3] $\cdot/1/\infty$: il y a un *serveur* (1) et un *buffer* d'entrée à capacité illimitée (∞). Le terme (.) correspond aux caractéristiques du temps de service demandé par un client au serveur. On note $(.) = D$, $(.) = M$, $(.) = GI$ ou $(.) = G$ suivant que la suite $\{\sigma(n), n \in \mathbb{N}\}$ des temps de service, est constante, i.i.d. exponentiellement distribuée, i.i.d. ou enfin stationnaire ergodique. On peut adjoindre à la notation de Kendall une information sur la politique de service considérée. Dans la suite, on considère toujours et de façon implicite une politique PAPS.

On a représenté ci-dessous, figure 1.6, quelques exemples élémentaires de files d'attente sous forme de graphes d'événements.
La transition t correspond au serveur, la place p au buffer d'entrée. On modélise un buffer à capacité limitée à l'aide d'une place supplémentaire p' comprenant autant de jetons qu'il y a de places dans le buffer. Le nombre de jetons dans la place rebouclant la transition t correspond au nombre de serveurs.

On peut maintenant considérer des réseaux de files d'attente. On a représenté en figure 1.7 quatre variations de réseaux à deux files d'attente en série.

[3] La notation complète est $\cdot/\cdot/1/\infty$. Le premier terme (.) correspond au processus d'arrivée.

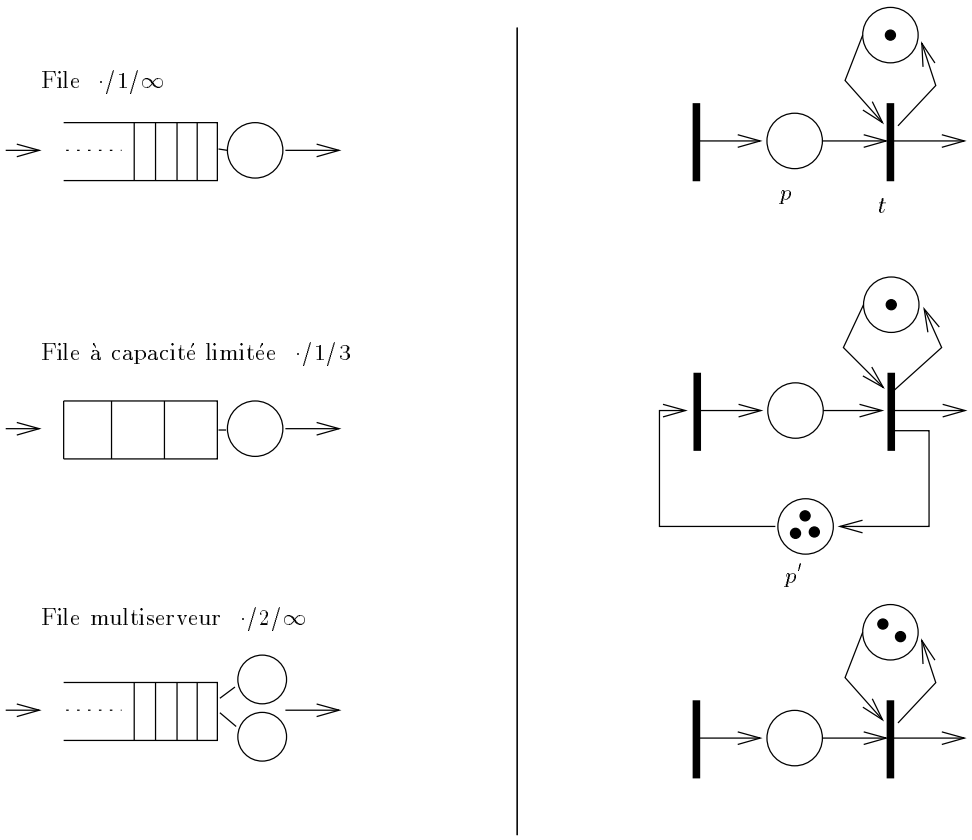


Figure 1.6: Représentation de files d'attente sous forme de graphes d'événements.

Dans les trois derniers exemples de la figure 1.7, la deuxième file est à capacité limitée, en conséquence certains clients peuvent se trouver bloqués. Décrivons plus précisément les mécanismes de blocage représentés en figure 1.7.

(A) Aucun blocage.

(B) Blocage avant service. Un service ne peut commencer en file 1 que si il y a une place de libre dans le buffer de la file 2.

(C) Blocage après service. Un client ayant terminé son service en file 1 et trouvant le buffer de la file 2 plein, doit attendre au niveau de la file 1 (son lieu d'attente est matérialisé par la place $\overline{p}$). Ce faisant, il interdit à un nouveau client d'être servi en file 1. Ce mécanisme est représenté par le circuit $(\overline{p}, \tilde{t}, \tilde{p}, t_1)$. Le mécanisme est le

(D) Blocage général (introduit par Cheng et Yao [38]). La file dispose de deux buffers, un en entrée et un en sortie. Il existe d'autre part une limitation sur le nombre total de clients présents dans la file. Une telle file est notée $\cdot/1/(a,b,c)$ où a, b et c sont respectivement les capacités des buffers d'entrée et de sortie et la capacité totale de la file (qui peut être différente de $a+b$). L'exemple de la figure 1.7.(D) correspond à une file $\cdot/1/(\infty,0,\infty)$ en

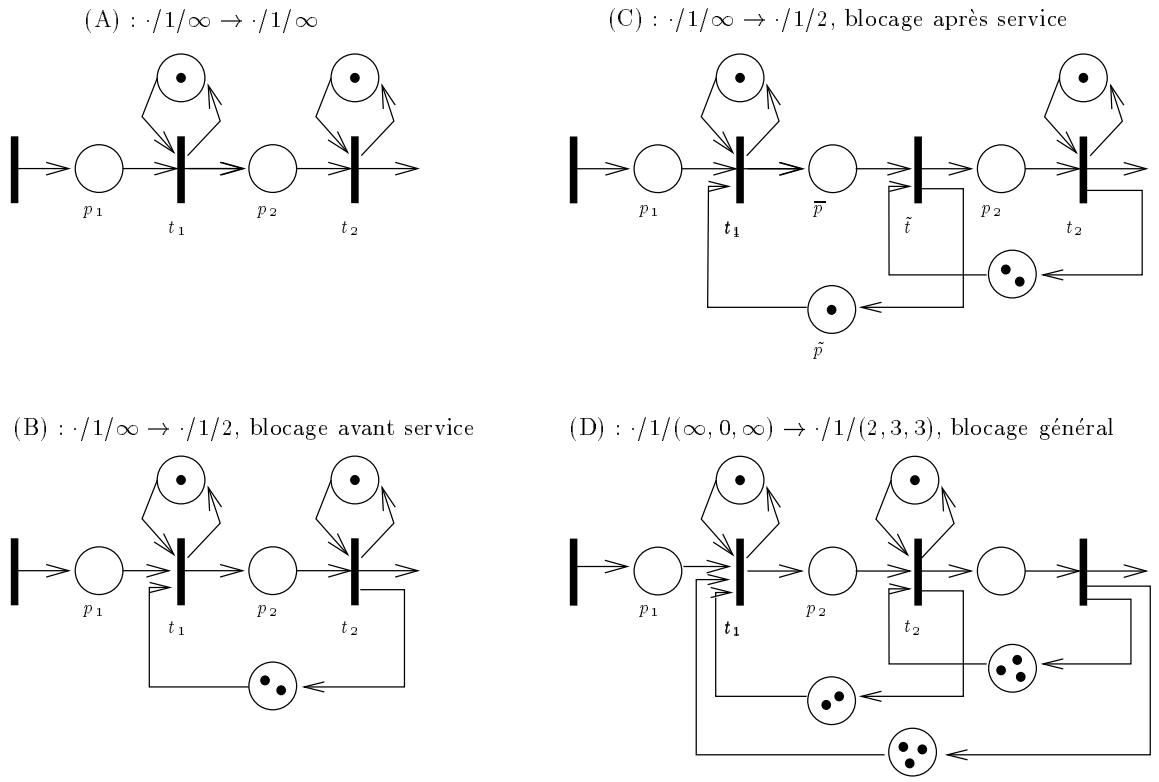


Figure 1.7: Files d'attente en série. Représentation sous forme de graphes d'événements.

tandem avec une file $\cdot/1/(2,3,3)$. Le cas $a = b = c$ correspond à une des variantes du célèbre mécanisme de *kanban*.

La notation de Kendall généralisée $\cdot/\cdot/(a,b,c)$ permet de définir complètement un réseau sans avoir à spécifier le mode de blocage.
Un mode de blocage qui ne peut être représenté à l'aide d'un graphe d'événements est le blocage avec "réémission" : un client se trouvant bloqué après service en file 1 recommence un nouveau service en file 1.

Dans les systèmes décrits en figures 1.6 et 1.7, la modélisation du flux d'arrivée de clients n'a pas été réalisée. On propose en figure 1.8, deux exemples de réseaux de files d'attente complètement spécifiés.

- Système ouvert, figure 1.8.I. On considère un flux d'arrivée de clients modélisé par la transition recyclée u. La suite des temps de tir de la transition u correspond aux interarrivées entre clients.
- Système fermé, figure 1.8.II. On considère un système autonome avec un nombre de clients constant. Un client terminant son service en file 2 est routé vers la file 1. Dans l'exemple considéré, figure 1.8.II, il y a 3 clients (nombre de jetons dans les places p_1 et p_2).

A partir des ingrédients des figures 1.6, 1.7 et 1.8, on peut construire de nombreuses variantes de

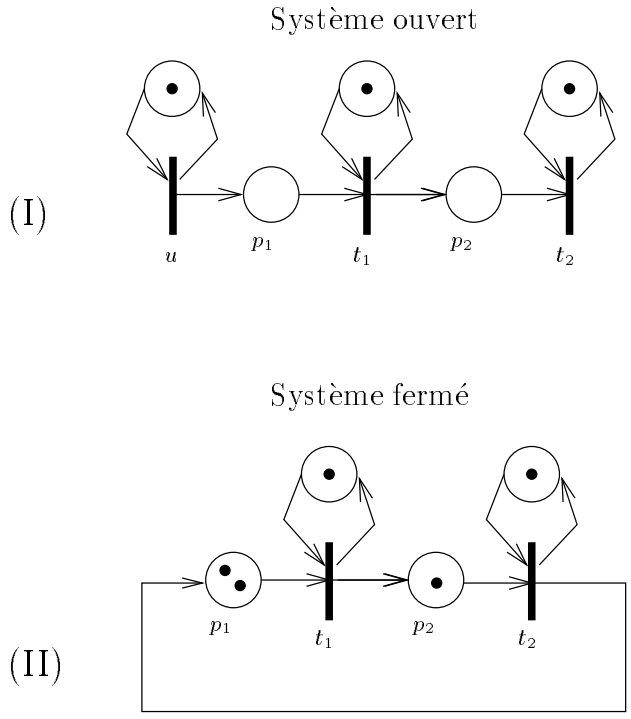


Figure 1.8: Files d'attente en série. Système ouvert et système fermé.

séries de files d'attente de longueur arbitraire. Le modèle de la figure 1.8.II avec k files $G/1/\infty$ sera spécifiquement étudié au chapitre 6.

On considère le réseau de la figure 1.8.II. Soit $\{\alpha(n), n \in \mathbb{N}\}$ et $\beta(n), n \in \mathbb{N}\}$, la suite des temps de service des serveurs 1 (t_1) et 2 (t_2) respectivement. L'évolution au plus tôt du réseau peut être modélisée par un système (max,+) de dimension 3 (la place p_1 doit être éclatée en deux places, voir figure 1.2). Les deux premiers dateurs correspondent aux transitions t_1 et t_2 et le troisième à la transition fictive ajoutée. On obtient :

$$x(n+1) = A_1(n) \otimes x(n),\ A_1(n) = \begin{pmatrix} \alpha(n) & -\infty & 0 \\ \alpha(n) & \beta(n) & -\infty \\ -\infty & \beta(n) & 0 \end{pmatrix} .$$

Les limites du premier ordre $\lim_n n/x_i(n)$ correspondent au débit du réseau. A partir des limites du second, on peut calculer des quantités telles que le nombre de clients par file ou le temps d'attente avant service. Ainsi le temps d'attente avant le n-ième service à la file 2 est donné par $z(n) = y_2(n) - (\ y_1(n-1) + \alpha(n-1)\ )$.

1.4 Automates temporisés

Un formalisme utilisé pour l'étude des SED et non encore mentionné est celui des automates et des langages formels. Cette approche a été introduite par Ramadge et Wonham [120].

Il s'agit d'un outil mathématique traitant des aspects purement "logiques" des SED. De façon très schématique, à chaque événement pouvant se produire dans le système est associé une *lettre*, celles-ci formant un *alphabet*. L'ensemble des comportements possibles ou souhaités du système correspond à un *langage* L sur cet alphabet. Le langage L est représenté sous la forme du langage reconnu par un automate (voir définition 1.4.1).
Des travaux récents ont consisté à temporiser cet objet logique. La notion d'automate temporisé, ou encore automate (max,+), que nous présentons ici a été proposée par Gaubert [69]. Un des exemples proposés, le modèle de stockage, est une adaptation de celui considéré dans [69]. Le second exemple, un réseau de Petri avec choix, est inspiré d'un travail en préparation [71]. Les systèmes tâches-ressources sont un autre exemple de modèles se représentant sous forme d'automates (max,+). Ils seront introduits et étudiés au chapitre 9.

Définition 1.4.1. *On appelle automate (déterministe) sur l'alphabet $\mathcal{A}$ un quadruplet*

$$(K, \delta, K_e, K_s) \ .$$

K est l'ensemble des états, $K_e \subset K$ l'ensemble des états d'entrée et $K_s \subset K$ celui des états de sortie. L'application partielle (i.e. non nécessairement partout définie) $\delta : \mathcal{A} \times K \to K$ est dite fonction de transition de l'automate.

Un automate est représenté sous forme d'un graphe à $k = |K|$ nœuds comportant trois types d'arcs, les arcs d'entrée, de sortie et les arcs internes, ces derniers étant valués. Un arc interne valué par $a \in \mathcal{A}$ relie les nœuds i et j si $\delta(a, i) = j$. On repère les états d'entrée par un arc entrant et les états de sortie par un arc sortant, voir figure 1.9.
A tout chemin dans le graphe, on associe un mot constitué par la suite des labels des arcs rencontrés. Un mot est dit *reconnu* si il correspond à un chemin menant d'un état d'entrée à un état de sortie. L'ensemble des mots reconnus forme le *langage reconnu* par l'automate.

Un automate (max,+) est défini comme un automate avec une temporisation ($\in \mathbb{R}^+$) associée à chaque arc interne. On note $A(a, i, j)$ la temporisation associée à l'arc $i \to j$ valué par a. Si l'on n'a pas $\delta(a, i) = j$, on pose $A(a, i, j) = -\infty$. On définit le vecteur ligne d'entrée α et le vecteur colonne de sortie β de la façon suivante : $\alpha_i = 0$ (resp. $\beta_i = 0$) si $i \in K_e$ (resp. $i \in K_s$) et $\alpha_i = -\infty$ (resp. $\beta_i = -\infty$) sinon.
Ceci permet d'associer une durée finie à chaque mot reconnu par l'automate. La durée d'un mot w est la temporisation maximale des chemins de label w. Au contraire, un mot non reconnu aura une durée $-\infty$. Cette durée est calculée de façon récursive.

On associe un dateur x_i à chaque état i. On note X la durée d'un mot. Soit w un mot et $a \in \mathcal{A}$ une lettre, on a :

$$\begin{cases} x_i(wa) & = \max_j\{x_j(w) + A(a, j, i)\}, \ \ x(\emptyset) = \alpha \\ X(wa) & = \max_j\{x_j(wa) + \beta_j\} \ . \end{cases}$$

Il apparaît maintenant clairement que ce modèle possède une structure (max,+) linéaire. A chaque lettre $a \in \mathcal{A}$, on associe une matrice (max,+) définie comme suit :

$$A(a) = (A(a, i, j), \ i, j \in \{1, \ldots, k\}) \ .$$

On obtient à partir de l'équation (1.4) et pour un mot $w = a_1 a_2 \dots a_n$:

$$X(w) = \alpha A(a_1) A(a_2) \cdots A(a_n) \beta \, .$$

La spécificité du modèle provient de ce que le produit de matrices (max,+) est contraint par un langage reconnaissable par un automate[4].

Un problème spécifique intéressant est celui de l'analyse de performance dans le pire des cas. On définit la durée maximale d'un mot reconnu de longueur $|w| = n$ par

$$\mathcal{T}(n) = \max_{w,\ |w|=n} X(w) = \alpha(\bigoplus_{w,\ |w|=n} x(w)) \beta = \alpha M^n \beta \, ,$$

où la matrice M est définie par $M = \bigoplus_{a \in \mathcal{A}} A(a)$, *i.e.* $M_{ij} = \max_a A(a,i,j)$.

Nous proposons maintenant une application puis une généralisation de ce modèle.

1.4.1 Un modèle de stockage

On considère un stock pouvant contenir 0, 1 ou 2 objets. On considère un alphabet à trois lettres correspondant aux événements suivants :

a : Un objet est ajouté au stock.

b : Deux objets sont retirés du stock.

c : Un objet est retiré du stock.

L'état 0 est l'état d'entrée et l'état 2 celui de sortie. Cela signifie que l'on part d'un stock vide et que l'on souhaite terminer avec un stock plein. On a représenté l'automate associé à ce modèle en figure 1.9. Les temporisations sont représentées entre parenthèses. Ainsi enlever deux objets du stock nécessite n_1 unités de temps.

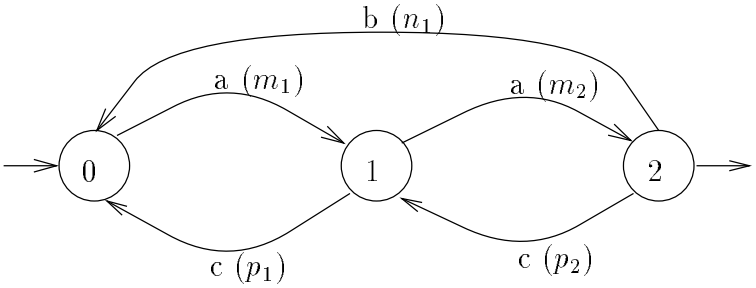


Figure 1.9: Automate temporisé associé à un modèle de stockage.

Le langage reconnu par l'automate est[5] : $L = (ac)^* a^2 \, (c(ca)^* a)^* \, \left(b(ac)^* a^2 \, (c(ca)^* a)^*\right)^*$.

[4] On remarquera que le produit de matrices est effectué de la gauche vers la droite, contrairement au cas des équations (1.3) et (1.9).

[5] On utilise la notation (classique) a^* pour le langage $\{\emptyset\} \cup \{a\} \cup \{a^2\} \cup \cdots$, la loi produit étant la concaténation, i.e. $a^2 = aa$.

Les matrices (max,+) associées au système sont :

$$A(a) = \begin{pmatrix} -\infty & m_1 & -\infty \\ -\infty & -\infty & m_2 \\ -\infty & -\infty & -\infty \end{pmatrix}, \; A(b) = \begin{pmatrix} -\infty & -\infty & -\infty \\ -\infty & -\infty & -\infty \\ n_1 & -\infty & -\infty \end{pmatrix}, \; A(c) = \begin{pmatrix} -\infty & -\infty & -\infty \\ p_1 & -\infty & -\infty \\ -\infty & p_2 & -\infty \end{pmatrix}.$$

1.4.2 Réseaux de Petri avec choix

Le modèle que nous présentons maintenant est une généralisation de l'automate (max,+) défini en section §1.4. On considère un modèle à deux niveaux.

- Niveau *logique* : on dispose d'un automate classique (i.e. non temporisé) sur l'alphabet $\mathcal{A}$. Cet automate définit un langage reconnaissable L.
- Niveau *temporel* : on dispose d'un ensemble de matrices $\{A(a), a \in \mathcal{A}\}$ à valeur dans $\mathbb{R}_{max}^{k\times k}$.

A chaque mot $w = a_1 \ldots a_n$ du langage L, on associe une matrice $A(w)$ définie comme suit :

$$A(w) = A(a_1) \otimes A(a_2) \cdots A(a_n) .$$

La généralisation provient de ce que les matrices $A(a)$ peuvent être quelconques (en particulier leur dimension k n'est pas à priori liée à la taille de l'espace d'état de l'automate).

Reprenons le modèle d'atelier flexible de la figure 1.4. Deux types de produits j_1 et j_2 sont fabriqués sur deux machines M_1 et M_2. Le routage des produits est $\sigma(j_1) = M_1M_2$ et $\sigma(j_2) = M_1$. Les temps d'exécution sont (α, β, γ) comme définis en §1.3.2. La différence avec l'atelier étudié en §1.3.2 provient de ce que l'ordonnancement des produits sur les machines est libre et non fixé. Un tel atelier peut se modéliser par le réseau de Petri (avec choix) de la figure 1.10.

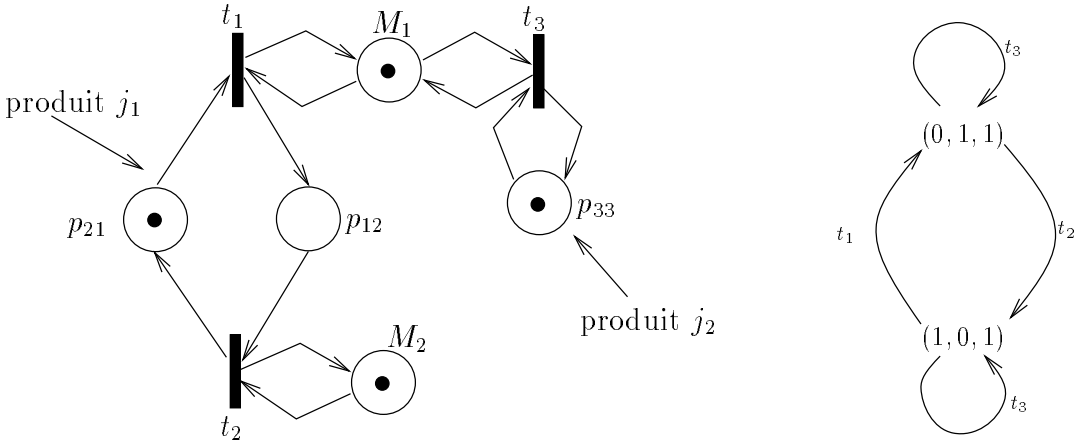


Figure 1.10: Atelier flexible. Modélisation sous forme de réseau de Petri. Graphe des marquages associés.

On a également représenté en figure 1.10, le graphe G des marquages associés à ce réseau de Petri. Les nœuds de G sont les marquages atteignables (pour une définition formelle, voir §8.1,

définition 8.1.2). On représente un arc valué par t_i entre le nœud M et le nœud M' si le tir de t_i fait passer du marquage M au marquage M'.
Dans l'exemple considéré, on représente le marquage par le triplet $(\mathcal{M}(p_{21}), \mathcal{M}(p_{12}), \mathcal{M}(p_{33}))$. Il n'y a que deux marquages atteignables, $(1,0,1)$ et $(0,1,1)$.
Ce graphe des marquages peut être vu comme un automate, définition 1.4.1. L'état initial correspond au marquage initial et tous les états sont des états finaux (on peut également spécifier un état final particulier). Soit L le langage reconnu par l'automate, pour l'exemple de la figure 1.10, on a $L = (t_3^* t_1 t_3^* t_2)^* \cup (t_3^* t_1 t_3^* t_2)^* t_1$. On définit un vecteur-ligne de dateurs :

$$x \ : \ L \to \mathbb{R}_+^4 \, .$$

Le dateur $x_1(w)$ correspond à la date de disponibilité du jeton du circuit $(t_1, p_{12}, t_2, p_{21})$ après tir des transitions dans l'ordre défini par w. Les dateurs $x_2(w), x_3(w)$ et $x_4(w)$ sont définis de la même façon pour les jetons des places p_{33}, M_1 et M_2 respectivement.
On définit les matrices :

$$A(t_1) = \begin{pmatrix} \alpha & -\infty & \alpha & -\infty \\ -\infty & 0 & -\infty & -\infty \\ \alpha & -\infty & \alpha & -\infty \\ -\infty & -\infty & -\infty & 0 \end{pmatrix}, \ A(t_2) = \begin{pmatrix} \beta & -\infty & -\infty & \beta \\ -\infty & 0 & -\infty & -\infty \\ -\infty & -\infty & 0 & -\infty \\ \beta & -\infty & -\infty & \beta \end{pmatrix},$$

$$A(t_3) = \begin{pmatrix} 0 & -\infty & -\infty & -\infty \\ -\infty & \gamma & \gamma & -\infty \\ -\infty & \gamma & \gamma & -\infty \\ -\infty & -\infty & -\infty & 0 \end{pmatrix}.$$

Le dateur $x(w)$ associé au mot $w = t_{i1} t_{i2} \cdots t_{in}$ est solution du système (max,+) linéaire :

$$x(w) = x(0) \otimes A(t_{i1}) A(t_{i2}) \cdots A(t_{in}) \, .$$

Cette modélisation s'applique à la classe générale des réseaux de Petri conservatifs, voir [71]. Toutefois, on ne peut pas modéliser toutes les politiques de résolution des conflits. Précisons ce dernier point.
Dans un réseau de Petri avec choix, il est nécessaire lorsqu'un jeton habilite plusieurs transitions de préciser la sémantique du tir (compétition ou pré-sélection). On parle de politique de résolution de conflits. La politique de compétition consiste à allouer le jeton à la transition dont le tir s'achève le premier. Dans le cas de la pré-sélection, une fonction de routage associée aux places décide de l'allocation des jetons. En général, les dépendances entre les tirs des transitions ne sont pas bornées (i.e. le n-ième tir de t_i peut être déclenché par le $(n-p)$-ième tir de t_j avec p arbitrairement grand) ce qui rend l'écriture d'équations de récurrence beaucoup plus difficile, voir [7]. L'approche classique consiste à se limiter au cas où les temps de tir des transitions sont exponentiellement distribués. En raison de la propriété sans mémoire de la loi exponentielle, on peut décrire l'évolution du système à l'aide d'une chaîne de Markov à temps continu sur l'espace des marquages atteignables, voir par exemple [1]. Si l'on souhaite considérer d'autres types de temps de tir, l'espace d'état doit incorporer les temps de tir résiduels des transitions ce qui rend

la modélisation plus difficile à opérer (explosion de l'espace d'état,...).

Dans notre modèle, l'ordre de déclenchement des événements n'est pas déterminé par des règles temporelles mais par une suite de décisions logiques prises au niveau de l'automate (non-temporisé) des marquages. Cette restriction permet de modéliser une politique de pré-sélection mais pas une politique de compétition (dans ce dernier cas, la résolution des conflits est faite au niveau temporel). En contrepartie, il devient possible de considérer n'importe quel type de loi pour les temps de tir des transitions sans modifier les dimensions du modèle. Pour le niveau logique, cette dimension est celle du graphe des marquages (comme dans l'approche markovienne classique), et pour le niveau temporel, la dimension des matrices (max,+) est égale au nombre de jetons dans le réseau de Petri.

1.5 Optimisation Dynamique

Cette section est inspiré de l'article de Yakovenko et Kontorer [142]. Dans ce papier, les auteurs font le point sur les systèmes (max,+) linéaires en économie et en théorie de la décision.

Rappelons brièvement ce qu'est un problème d'optimisation dynamique en temps discret et à espace d'état fini. On considère un espace d'état fini $\mathcal{K}$. La dynamique du système est décrite par un ensemble de chemins ou trajectoires. Une trajectoire est une suite d'états dans l'espace $\mathcal{K}$. Soit une économie composée de k biens consommables. On décrit le niveau de stock en biens par un élément de $\mathbb{N}^k$. Si on suppose qu'il existe un stock maximal pour chaque bien, on obtient un espace d'état fini $\mathcal{K} \subset \mathbb{N}^k$. Une trajectoire correspond à l'évolution temporelle des quantités de biens disponibles. Une fonction de transition $a : \mathcal{K} \times \mathcal{K} \to \mathbb{R} \cup \{-\infty\}$ est associée au passage d'un état à un autre. Certaines transitions peuvent être interdites, ce qui est spécifié en donnant la valeur $-\infty$ à $a(.,.)$. Cette fonction est souvent appelée *fonction de coût* en contrôle et *utilité* en économie. En économie, l'interdiction de certaines transitions correspond souvent à des limitations en terme de capacité de production.
En théorie du contrôle, l'objectif est de déterminer une trajectoire optimale, c'est-à-dire minimisant le coût. En économie classique, sous une hypothèse dite d'*information parfaite*, on considère que l'économie évolue de façon rationnelle. Cela signifie que la trajectoire réelle est effectivement celle qui maximise l'utilité. Le choix significatif devient celui du choix de l'état final étant donné l'état initial. Dans les deux cas, il est essentiel de déterminer les trajectoires optimales.

Le critère à optimiser en l'absence de *taux d'actualisation*, resp. avec un taux d'actualisation r, est :

$$\sum_n a(x(n), x(n+1)), \text{ resp. } \sum_n r^n \times a(x(n), x(n+1)) ,$$

où $\mathbf{x} = \{x(0), x(1), \ldots, x(n), \ldots\}$ est la trajectoire. On se restreint par la suite au critère sans actualisation. On ajoute souvent au problème un gain[6] terminal. Le gain maximal en horizon

[6]Par la suite, on considère des gains plutôt que des coûts de façon à travailler avec la fonction max plutôt que la fonction min.

N (souvent appelé fonction valeur) est alors obtenu comme solution de l'équation :

$$v(x(0)) = \max_{\mathbf{x}} \sum_{n=0}^{N-1} a(x(n), x(n+1)) + g(x(N)) \,, \tag{1.10}$$

où la fonction g correspond au gain terminal. On introduit la matrice A définit par $A_{ij} = a(i,j)$ et on identifie la fonction $g : \mathcal{K} \to \mathbb{R}$ à un vecteur colonne. L'équation (1.10) peut se récrire sous la forme de (1.2) :

$$v(x(0)) = (A^N \otimes g)_{x(0)}, \quad v = A^N \otimes g \,. \tag{1.11}$$

La valeur propre de la matrice A, voir §1.2, correspond au gain (resp. à l'utilité) moyen(ne) le long d'une trajectoire optimale.

En général, en optimisation, on s'intéresse plus à la forme des trajectoires optimales qu'à la valeur du gain moyen. Ces trajectoires sont directement reliées aux vecteurs propres de l'équation spectrale (1.4), plus connue en optimisation sous le nom d'équation de Bellman stationnaire. Essayons d'expliquer cela.

Lorsque l'on résout l'équation (1.10) en horizon fini, on se heurte au problème de l'existence d'effets de bord. On peut éliminer ces effets en choisissant une fonction g adaptée. On choisit g de façon à ce que la solution de (1.11) soit indépendante de l'horizon N. Plus précisément, la trajectoire optimale du problème (1.11) sous l'horizon N doit constituer les N premières étapes de la trajectoire optimale sous l'horizon N' pour $N' > N$.
Les fonctions g vérifiant cette propriété sont les solutions de l'équation de Bellman, c'est-à-dire les vecteurs propres de la matrice (max,+) A.

Un autre problème classique est celui de l'optimisation en horizon infini. Soit g un vecteur propre de la matrice A. D'après ce que l'on vient de voir, on peut associer à g une trajectoire optimale indépendante de l'horizon N et donc, par passage à la limite, une trajectoire infinie optimale. Notons qu'il n'y a pas unicité de la trajectoire optimale, une matrice (max,+) pouvant avoir plusieurs vecteurs propres. Ce phénomène a été très étudié en économie mathématique sous le nom de *turnpike theory*, voir McKenzie [108]. En particulier il existe une importante littérature définissant des critères permettant de choisir entre trajectoires infinies optimales, voir [142] et ses références.

Il est également classique de considérer une fonction de transition $a_n(.,.)$ variable au cours du temps. En économie, les variations peuvent correspondre à des changements technologiques, à des variations de la taille ou de la composition de la population voir à des modifications des goûts et des comportements. La modélisation doit alors se faire à l'aide de matrices (max,+) variables. L'équation (1.11) devient donc :

$$v = A(1)A(2)\cdots A(N) \otimes g \,. \tag{1.12}$$

Pour pouvoir calculer v de façon récursive, il est nécessaire d'introduire les variables $v_k = A(k)A(k+1)\cdots A(N) \otimes g$. On obtient :

$$v_k = A(k)v_{k+1} \text{ et } v_0 = v \,. \tag{1.13}$$

L'équation rétrograde (1.13) est connue sous le nom d'équation de la *programmation dynamique*.

1.6 Mécanique Statistique

Le modèle de Frenkel-Kontorova est étudié en mécanisme statistique. Ce modèle simplifié permet d'appréhender les phénomènes de transitions de phase apparaissant dans l'étude des structures cristallines et du magnétisme des métaux rares. Pour une présentation plus détaillée, on se reportera à Chou et Griffiths [41] et surtout à l'article de synthèse de Griffiths [80].

Le modèle de Frenkel-Kontorova peut être décrit comme un système de particules reliées par des ressorts et placées dans un potentiel périodique, voir figure 1.11.

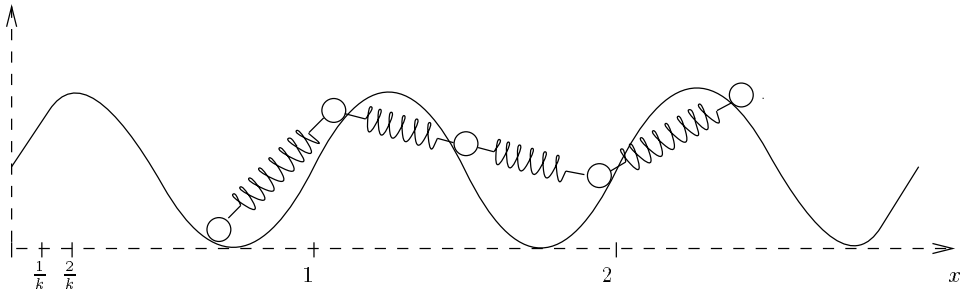


Figure 1.11: Modèle de Frenkel-Kontorova.

L'énergie potentielle totale de ce système est de la forme :

$$H = \sum_n H(x_n, x_{n+1}) = \sum_n \left(W(x_{n+1} - x_n) + V(x_n)\right) \ , \tag{1.14}$$

où x_n est la position de la n-ième particule, V est un potentiel périodique de période 1 et W est l'énergie potentielle du ressort. On a le plus souvent $V(u) = K\cos(2\pi u)$ et $W(u) = 1/2(u-\gamma)^2$ où γ est la longueur du ressort au repos.

L'approche physique classique pour étudier un tel système est de rechercher les solutions de l'équation à l'équilibre $\partial H/\partial x_n = 0$. La résolution de cette équation est dans le cas présent particulièrement difficile. On propose ici une analyse sensiblement différente.

Etant donné une énergie de la forme (1.14), une procédure standard en mécanique statistique consiste à introduire l'opérateur de transfert F défini ci-dessous :

$$F(x, x') = \exp\left(-\frac{1}{T}H(x, x')\right) \ , \tag{1.15}$$

où T est la température. Les propriétés du système original sont étudiées à travers la limite de l'opérateur de transfert lorsque la température tend vers 0.

On considère dans la suite le modèle de Frenkel-Kontorova discrétisé avec une grille de résolution de pas $1/k$. Ainsi l'intervalle unité est-il découpé en $\{1/k, 2/k, \ldots, 1\}$. L'opérateur F se décrit comme une matrice positive de dimension $k \times k$ dont les coordonnées sont $F_{ij} = \exp(\ -1/T\ \tilde{H}_{ij}\ )$, où la matrice $\tilde{H}$ est définie par :

$$\begin{cases} \tilde{H}_{ij} = H(i/k, j/k) = W(j/k - i/k) + V(i/k) & \text{si } \ i \leqslant j \\ \tilde{H}_{ij} = H(i/k, j/k + 1) = W(1 + j/k - i/k) + V(i/k) & \text{si } \ i > j\ . \end{cases}$$

Les propriétés thermodynamiques du système sont déterminées par la plus grande valeur propre (dans l'algèbre usuelle) de F. Par le théorème de Perron Frobenius, celle-ci est réelle positive ainsi que son vecteur propre associé. On écrit la valeur propre sous la forme $\exp(-1/T\ \lambda)$ et le vecteur propre sous la forme $\exp(-1/T\ x) = (\ \exp(-1/T\ x_1), \ldots, \exp(-1/T\ x_k)\ )'$. On a :

$$\exp\left(-1/T\ \tilde{H}\right) \exp\ (-1/T\ x) = \exp\ (-1/T\ \lambda) \exp\ (-1/T\ x)\ . \tag{1.16}$$

Lorsque la température tend vers 0, on peut simplifier l'équation (1.16) en ne conservant que les termes dominants. On obtient alors l'équation (pour la ligne i) :

$$\begin{aligned} \max_j\ \exp\{(-\frac{1}{T}\tilde{H}_{ij} + x_j)\} \quad &\simeq \quad \exp\{-\frac{1}{T}(\lambda + x_i)\} \\ \Rightarrow \min_j\ (\ \tilde{H}_{ij} + x_j\ ) \quad &= \quad \lambda + x_i\ . \end{aligned} \tag{1.17}$$

L'équation (1.17) obtenue après passage au logarithme est une équation spectrale dans $\mathbb{R}_{min}$ en tout point similaire à celle définie en (1.4). On peut bien sûr se ramener à un système (max,+) linéaire en considérant $-\tilde{H}$, $-\lambda$ et $-x$.

Il existe une autre méthode, dans l'esprit de la programmation dynamique, voir §1.5, pour aboutir à l'équation (1.17). On commence par fixer la position x_1 de la première particule dans le potentiel, puis on détermine la position $x_2 \geqslant x_1$ de la deuxième particule de façon à minimiser l'énergie H pour l'ensemble des deux particules. On procède ensuite par récurrence. On note $h(n)_i, i = 1, \ldots, k$, l'énergie minimale d'un système à n particules dont la plus à gauche est en position i/k. En particulier, on a $h(1)_i = V(i/k)$. On obtient :

$$h(2)_i = \min_{j \geqslant i}\ (\ W(i/k - j/k) + V(i/k) + V(j/k)\ ) = \min_j (\ \tilde{H}_{ij} + h(1)_j\ )\ . \tag{1.18}$$

Par récurrence et en utilisant les notations de l'algèbre (min,+), on obtient :

$$h(n+1) = \tilde{H} \otimes h(n) = \tilde{H}^n \otimes h(1)\ . \tag{1.19}$$

L'équation (1.17) est l'équation spectrale associée à (1.19).

La valeur propre λ de H est *l'enthalpie* moyenne par particule. On s'intéresse également aux configurations de particules minimisant l'énergie H. Soit h un vecteur propre de $\tilde{H}$. On définit une fonction τ où $\tau(i)$ est un des indices réalisant le minimum de $(\tilde{H}_{ij} + h_j)$. A partir des itérés de cette fonction, on obtient une configuration *périodique* d'énergie minimale.

La densité spatiale de particules dans une configuration minimale est le paramètre physique le plus étudié. Pour une configuration périodique, cette densité est un rationnel. Le phénomène de *transition de phase* consiste en l'existence de discontinuités de cette densité en fonction des paramètres (γ, K) du modèle. D'un point de vue algébrique, cela s'interprète comme la discontinuité de la cyclicité (voir chapitre 3) en fonction des coordonnées de la matrice. De très beaux graphiques illustrant le phénomène sont proposés dans [80].

Un système est à interaction de type *plus proche voisin* si son énergie potentielle est de la forme $H = \sum H(x_n, x_{n+1})$ (généralisation de (1.14)). L'approche présentée ici s'applique à de tels systèmes à condition de pouvoir borner uniformément la distance entre deux particules de façon à représenter le système par une matrice de taille finie.

1.7 Conclusion

Les exemples proposés dans ce chapitre sont loin d'épuiser l'ensemble des problèmes et modèles représentables sous forme (max,+) linéaire.
Parmi les modèles que l'on a pas présentés, le plus célèbre et le plus ancien est celui de la recherche du plus long (ou court) chemin dans un graphe valué. Il s'agit d'une application directe de l'interprétation *algèbre de chemin* décrite en §1.2. Cette application est connue depuis la fin des années 50, voir les références de [50], chapitre 1. De nombreux problèmes en recherche opérationnelle peuvent être abordés de cette façon, voir [78] pour une présentation détaillée. L'ordonnancement au plus tôt sans contraintes de ressources est une instance du même problème et sera considéré au chapitre 4. Le problème plus compliqué de l'ordonnancement avec contraintes de ressources peut également être abordé par une approche (max,+) linéaire [71].
On peut également représenter sous forme (max,+) linéaire des réseaux de communication, voir [8] ou Braker [28] pour le réseau ferroviaire hollandais.
D'autre types de réseaux de files d'attente que ceux présentés en §1.3.3 peuvent être considérés. Parmi ceux-ci, on peut mentionner les réseaux série-parallèle, les modèles de rendez-vous et enfin des réseaux avec relations de précédence entre clients.
Un autre domaine très riche et non abordé dans ce chapitre (et dans cette thèse) est celui des systèmes (max,+) linéaires continus. On considère l'espace $\mathcal{F}$ des fonctions d'un ensemble E dans $\mathbb{R} \cup \{-\infty\}$. On définit l'opérateur $A : \mathcal{F} \to \mathcal{F}$ par :

$$Af(x) = \sup_{y \in E} \ (a(x,y) + f(y)) \stackrel{\text{def}}{=} \int\!\!\!\!\!\!\bigoplus_{E} a(x,y) \otimes f(y) \ .$$

On parle parfois d'intégrale de Maslov de f par rapport à la "mesure" a. Il existe une théorie complète de l'intégration de Maslov. Parmi les domaines d'application, on peut mentionner l'étude des solutions de certaines équations aux dérivées partielles (équation de Schrödinger, de Hamilton-Jacobi). Pour de plus amples détails, on se reportera au volume [107].

Chapitre 2

Minimal Representation of Uniform Recurrence Equations

Représentation Minimale d'Équations Récurrentes Uniformes

Dans ce chapitre, nous étudions des systèmes d'Équations Récurrentes Uniformes. L'objectif est de minimiser le nombre de cases mémoires nécessaires pour mener le calcul d'un tel système. Le modèle étudié est très général et se situe en amont des modèles étudiés par la suite. De fait, on verra en section 2.7 que les systèmes (max,+) linéaires sont un cas particulier d'Équations Récurrentes Uniformes. Une application des résultats de ce chapitre est la minimisation de la taille du système (max,+) linéaire représentant un graphe d'événement donné.

Ce chapitre provient d'un travail commun avec Bruno Gaujal et Alain Jean-Marie [74]. Une partie des résultats a été présentée dans [73]. Les auteurs tiennent à remercier tout particulièrement Jean-Claude Bermond, Alain Darte, Mike Robson et Stéphane Gaubert pour de nombreuses discussions.

We consider a system of uniform recurrence equations of dimension one. We show how the computation can be carried using minimal memory size with several synchronous processors. This result has applications in register minimization for digital circuitry and parallel simulation of timed event graphs.

2.1 Introduction

The model under study will be the Uniform Recurrence Equations (URE) [93].

Definition 2.1.1 (URE). *We consider E-valued variables $X_i(n), i \in V = \{1, \ldots, k\}, n \in K$, where E is an arbitrary set and $K \subset \mathbb{Z}^p$ for some $p \in \mathbb{N}$. These variables satisfy the equations*

$$X_i(n) = F_i(X_j(n - \gamma)), (j, \gamma) \in \mathcal{D}_i, \forall n \in K \; . \tag{2.1}$$

The sets $\mathcal{D}_i$ are finite non empty subsets of $\{1, \ldots, k\} \times \mathbb{Z}^p$.

The integers γ are called the *delays*. The system $\mathcal{S}$ defined by Equation (2.1) is said to be uniform because the dependence sets $\mathcal{D}_i$ do not depend on n. Note that it is possible to have two delays $\gamma, \gamma' \in \mathbb{Z}^p, \gamma \neq \gamma'$ such that $(j, \gamma) \in \mathcal{D}_i$ and $(j, \gamma') \in \mathcal{D}_i$. There is no restriction on the generality of the functions F_i considered.

There are various motivations to study URE. They appear in the description of differential equations using finite difference methods or in the study of discrete event systems. The case $p > 1, K = \mathbb{Z}^p$ has often been studied in the literature, see [93]. In such a case, some of the major issues are the constructivity [93] and loop parallelization [53]. These problems and others appearing in this framework will be discussed in §2.3.1.

In this paper, we consider only the simple case where $K = \mathbb{Z}$ (systems of dimension one). We investigate the problem of minimizing the number of "memory locations": we want to determine the minimal memory size that is needed to compute all the variables $X_i(n)$ of Equation (2.1) using parallel processors with a shared memory.

We show that the solution of this problem has many applications. It can be used in order to obtain the most efficient representation of the system for simulation purposes. This aspect of the problem will be investigated in §2.7. In a quite different context, URE appear in the modeling of logical circuits, systolic arrays or program loops. The minimization for URE enables us to obtain an optimal design of such circuits (in terms of number of registers). This application will be discussed in §2.6.

The paper is organized as follows. In Section 2.2, we precise the definition of a system of URE and we present two associated graphs, the *dependence graph* and the *reduced graph*. In Section 2.3 we describe the problem that we are going to address. Section 2.4 investigates the relations that can be found between *cuts* in the dependence graph and the memory size required for an execution of the URE; section 2.5 presents the interpretation of the above quantities in the reduced graph. Finally in Sections 2.6 and 2.7, two applications are described, for digital circuits and (max,+) linear systems respectively.

2.2 Basic Models

From now on, we consider URE of dimension 1. More precisely, we consider the set of variables $X_i(n), i \in V = \{1, \ldots, k\}, n \in \mathbb{Z}$ and the equations

$$X_i(n) = F_i(X_j(n-\gamma)), (j,\gamma) \in \mathcal{D}_i,\ n \in \mathbb{Z}\ , \tag{2.2}$$

where the sets $\mathcal{D}_i$ are finite non empty subsets of $\{1, \ldots, k\} \times \mathbb{N}$.

A system of URE $\mathcal{S}$ is *constructive* if given the values of the "negative" variables $X_i(n), n \leqslant 0$, involved in $\mathcal{S}$ (*initial data*), there exists an ordering of the equations such that, $\forall i, \forall n > 0$, all the variables present in the right hand side of the equation defining $X_i(n)$ are either "negative" variables or can be computed before $X_i(n)$. This condition is satisfied if and only if there is no cycle in the dependences, i.e. there does not exist $i_1, \ldots, i_p, i_{p+1} = i_1$ such that $(i_2, 0) \in \mathcal{D}_{i_1}, (i_3, 0) \in \mathcal{D}_{i_2}, \ldots, (i_1, 0) \in \mathcal{D}_{i_p}$.

Remark 2.2.1. We could have considered an apparently more general definition of URE allowing the delays γ to be negative in Equation (2.1.1). In this case, the constructivity assumption becomes
For each cycle $(i_1, \gamma_1), \ldots, (i_p, \gamma_p), i_{p+1} = i_1$ such that $(i_{j+1}, \gamma_{j+1}) \in \mathcal{D}_{i_j}$, $j \in \{1, \ldots, p\}$ then $\sum_{j=1}^{p} \gamma_j > 0$. Under the constructivity assumption, it is possible to come back to Equation (2.1.1) through a simple renumbering of the variables (i.e. $X_i(n) := X_i(n + c_i)$ for some constant $c_i \in \mathbb{Z}$ independent of n).

From now on, the system $\mathcal{S}$ that we consider is always assumed to be constructive. We are going to present two equivalent ways to describe $\mathcal{S}$: the dependence graph and the reduced graph.

Example 2.2.2. The illustrative examples to be presented in the following correspond to the system :

$$\left\{ \begin{array}{lcl} X_1(n) & = & F_1(X_3(n-1)) \\ X_2(n) & = & F_2(X_1(n-2)) \\ X_3(n) & = & F_3(X_2(n), X_4(n-2)) \\ X_4(n) & = & F_4(X_3(n-1), X_4(n-1)), \end{array} \right. \tag{2.3}$$

2.2.1 Dependence graph

We introduce the graph $\mathcal{D}$ of the dependences between the variables $X_i(n)$.

Definition 2.2.3 (Dependence graph). *The dependence graph associated with a system of URE is the graph $\mathcal{D}$ with $(V \times \mathbb{Z})$ as the set of nodes. There is an arc from the node (i, n) to the node (j, m) if $X_j(m) = F_j(X_i(n), \ldots)$ or equivalently if $(i, m-n) \in \mathcal{D}_j$ (notation: $(i, n) \to (j, m)$).*

The n-th *level* in $\mathcal{D}$ is the set of nodes $\{(i, n), 1 \leqslant i \leqslant k\}$. The i-th *column* in $\mathcal{D}$ is the set of nodes $\{(i, n), n \in \mathbb{Z}\}$. In the following, we will refer to nodes (i, n), $n \leqslant 0$ as *negative* nodes and nodes (i, n), $n > 0$ as *positive* nodes.

It is immediate from the definition of an URE that $\mathcal{D}$ is 1-periodic, i.e.

$$(i,n) \to (j,m) \Rightarrow (i,n+1) \to (j,m+1)\ .$$

Note also that the constructivity assumption implies that the graph $\mathcal{D}$ is acyclic.
We have represented in Figure 2.1 the dependence graph corresponding to the system of Example 2.2.2.

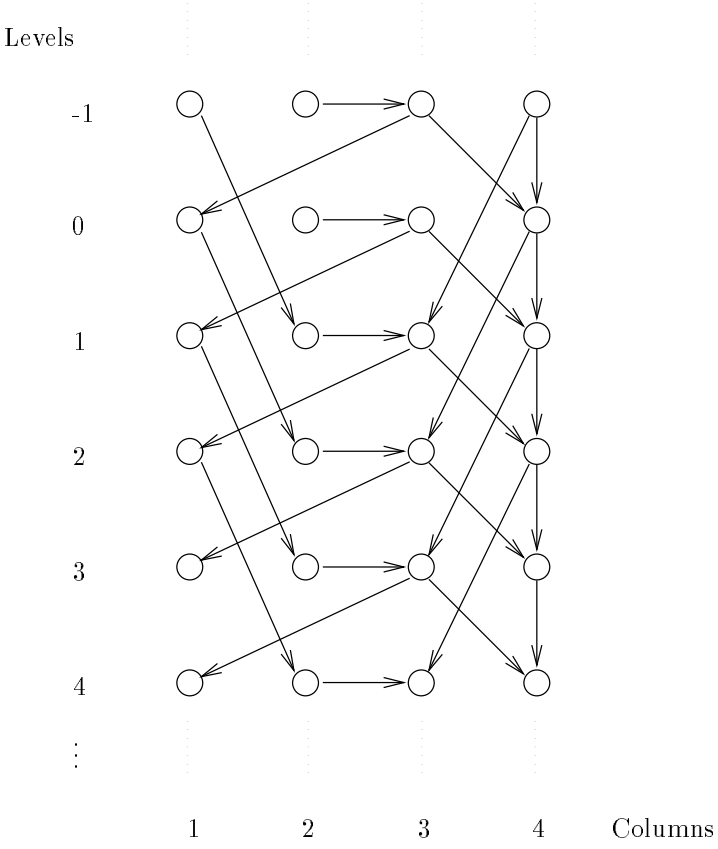


Figure 2.1: Dependence graph associated with the system $\mathcal{S}$ of Equation (2.3).

The dependence graph appears under various forms and names in the literature. For example, we can mention the following names: developed graph, task graph, PERT graph, unfolded process graph or activity network.

2.2.2 Reduced graph

Since the dependence graph is 1-periodic, it can be folded into a more compact form. This is how we construct the *reduced graph* $\mathcal{R}$ associated with the system $\mathcal{S}$.

Definition 2.2.4 (Reduced graph).
The reduced graph is an arc valued graph $\mathcal{R} = (V, E, \Gamma)$. *The set of nodes is* $V = \{1, \cdots, k\}$. *There is an oriented arc in* E *from* i *to* j *if*

$$\exists \gamma \in \mathbb{N} \ \ s.t.\ \ (i,\gamma) \in \mathcal{D}_j\ . \tag{2.4}$$

This arc is valued with the delay γ. If there exist several delays γ verifying condition (2.4), E contains several arcs between the nodes i and j, with corresponding values.

There is an arc from i to j in E, if and only if there are arcs from the column $(i,.)$ to the column $(j,.)$ in the dependence graph. Note that system $\mathcal{S}$ is constructive if and only if the sum of the delays along any circuit in $\mathcal{R}$ is positive.
The reduced graph associated with the system $\mathcal{S}$ of Example 2.2.2 is represented in Figure 2.2. The delays γ associated with the arcs are depicted in boxes.

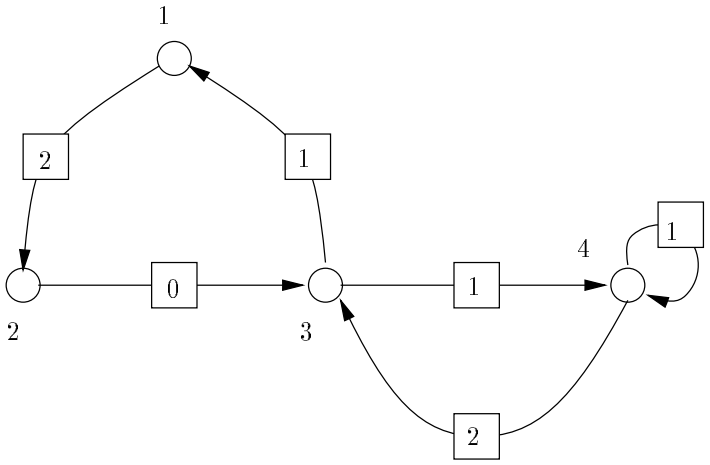


Figure 2.2: Reduced graph associated with the system $\mathcal{S}$ of Equation (2.3).

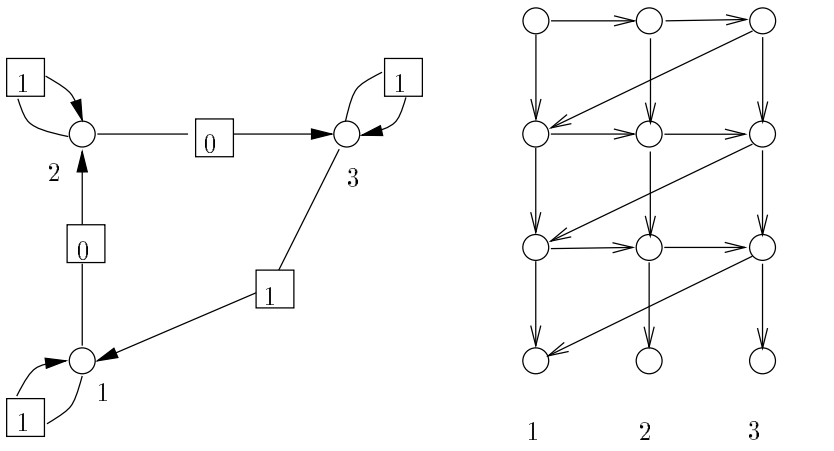
Reduced graphs appear in the literature under the following names : computation graph, Synchronous Data Flow (SDF) graphs, process graphs or uniform graphs.

2.2.3 Recycled case

In the following, we will particularly study a special case of URE, where the computation of the variable $X_i(n)$ cannot be done before the computation of $X_i(n-1)$. This case appears naturally in marked graphs to impose a FIFO behavior (see section 2.7) and in other applications. This constraint can be modeled by imposing a dependence between $X_i(n-1)$ and $X_i(n)$, for all i and n. Formally, it results in having $(i,1) \in \mathcal{D}_i, \forall i$, for the system of URE. Equivalently, it results in having a self loop with delay one (hence the name recycled) in each node of $\mathcal{R}$, or in having arcs between the nodes (i,n) and $(i,n+1)$ in $\mathcal{D}$.
Figure 2.3 depicts an example of a recycled system.

2.2.4 Preliminary remarks

It should be clear from the definitions that there is a one to one correspondence between the three models. Indeed, a system can be given by its reduced graph as well as its dependence graph or system of equations.
If $\mathcal{R}$ is not connected, then the system of URE is made of two or more independent systems which can be studied independently. In the rest of the paper we will always assume that the graph $\mathcal{R}$ is connected.



Recycled reduced graph Recycled dependence graph

Figure 2.3: Recycled reduced and dependence graphs.

In the following, we will study more precisely the relations between $\mathcal{D}$ and $\mathcal{R}$.

2.3 Parallel Executions

In this part, we will define the problem to be investigated in the rest of the paper. Roughly speaking, we want to minimize the memory size required to compute a system $\mathcal{S}$. In §2.3.1, we discuss the main issues that have been studied in the field of uniform recurrence equations and we explain the difference with the aim of this paper. Section §2.3.2 gives a formal definition of the question we investigate in the following.

2.3.1 Relations with the basic scheduling problem

Organizing efficient computations for uniform recurrence equations on parallel computers has now taken a considerable importance in the literature. In the past, the investigations have often been oriented towards speeding up the execution with no or little consideration for memory requirements.

Assume that initially the negative variables $X_i(n), n \leqslant 0$ are known. Assume also that each computation of a variable requires a time equal to 1. In a system of URE, the date at which the variable $X_i(n), n > 0$ will be computed is necessarily larger than the length of a longest path in $\mathcal{D}$ from level 0 to (i, n). A computation of a system of URE is *as soon as possible* if the time it takes to compute the variable $X_i(n)$ is exactly given by the longest path from level 0 to (i, n). A first question that can be addressed is :

What is the number of processors required to carry out
a computation as soon as possible Γ

This number is often called the *degree of parallelism* of the URE. In general, the solution is given by the size of the maximal anti-clique in the dependence graph.

Once this question is settled and provided a sufficient number of processors are available (*i.e.* larger than the degree of parallelism of the system), a problem is to characterize the as soon as possible schedule. This problem is often called the basic scheduling problem, see also Chapter 4. It has been proved that for the as soon as possible schedule, the time at which the variable $X_i(n)$ is computed is of the form $\lambda n + d_i$, where λ is called the cycle time of the system (see [42] [44]). Such a schedule is said to be linear. For systems of higher dimension (i.e. when $K = \mathbb{Z}^p$, $p > 1$ in Equation (2.1)), partial results on the optimality of linear schedules can be found in [52, 53].
When the number of available processors is fixed and less than the degree of parallelism of the URE, finding the optimal schedule becomes NP-hard, see [85].

All the results mentioned above are related with the problem of minimizing the number of processors used. This paper is concerned with the following dual problem : how much memory is necessary to carry out the computations of an URE, the number of processors being unlimited. First, we should say that, in general, a computation as soon as possible requires a lot of memory. It may not even be bounded in n when the reduced graph $\mathcal{R}$ is connected but not strongly connected. This makes the alternative to find a computation using a smaller memory size attractive. Second, the usual time-space trade-off tells us that some interesting results can be expected to arise when minimizing the memory size.

As one can expect, the schedule we will propose will not be as soon as possible in general. The time required to carry out the computation of variable $X_i(n)$ will be of the form $\lambda' n + d'_i$, where $\lambda' \geqslant \lambda$. Furthermore, the number of processors needed to carry out the computation will be greater than the degree of parallelism, but smaller than the memory size involved in the computation. In spite of these two drawbacks, the schedule we propose provides a new insight on the best ways to compute a system of URE and has interesting practical applications, see §2.6 and §2.7.
As shown in the following, the minimal size of the memory is related with *cuts* in the dependence graph (instead of anti-cliques for the minimal number of processors).

2.3.2 Pebble game

Let us work with an URE, $\mathcal{S}$ and its associated dependence graph $\mathcal{D}$ as defined in §2.2.1. We want to compute iteratively all the variables $X_i(n)$. At each step, the variables which are necessary to carry out the computations have to be stored in some memory locations.

We want to determine the minimal number of memory locations needed to compute all the variables $X_i(n)$.

We give a description of this problem in terms of a pebble game (see [119]).

Game 1. *We consider a dependence graph $\mathcal{D}$. At step 0, one puts a finite number of pebbles on negative nodes, i.e. on nodes $(i, n), n \leqslant 0$. By convention, we assume that at least one of these pebbles is on level 0. At each step of the game, the following moves are allowed.*

Move 1 *: Put a pebble on a node (i, n) if on each infinite oriented path (see Definition 2.4.1) ending in (i, n) there is at least one node with a pebble.*

Move 2 *: Remove a pebble from a node.*

An *execution* of the game is *successful* if all positive nodes receive a pebble during the execution. In the following, we will always refer to successful executions simply as executions. The set of (successful) executions will be denoted $\mathcal{E}$. Several variants for the number of moves allowed during a step are to be considered later on. Accordingly, a step of the game may have a duration different from 1 unit of time, see the discussion in Remark 2.3.3.

Definition 2.3.1 (configuration). *In an execution, the position of the pebbles at step $t \in \mathbb{N}$, is called the t-th configuration and is denoted $\mathcal{A}(t)$. In particular, $\mathcal{A}(0)$ corresponds to the set of initial pebbles.*

Pebbles correspond to memory locations. A pebble put on a node corresponds to the computation of a new data and its storage in the memory. The removal of a pebble corresponds to the erasing of a data from the memory. An execution corresponds to a computation of all the nodes in the dependence graph. Our objective will be to find an execution of the game using a number of pebbles which is as small as possible. The total number of pebbles used by an execution $e \in \mathcal{E}$ is

$$\mathcal{P}(e) \stackrel{\text{def}}{=} \max_{t \in \mathbb{N}} |\mathcal{A}(t)| \,,$$

where $|\mathcal{A}(t)|$ represents the cardinal of $\mathcal{A}(t)$.

Remark 2.3.2. Note that this definition of $\mathcal{P}(e)$ considers only the number of pebbles at the end of the step and not in intermediate stages (after move 1 and before move 2 for example). It corresponds to the assumption that both moves can be performed simultaneously.

Our objective can be redefined as follows:

Problem 1. *We want to find an execution $e_o \in \mathcal{E}$ such that $\mathcal{P}(e_o) = \min_{e \in \mathcal{E}} \mathcal{P}(e)$.*

The definition of Move 1 implies that we are allowed to perform function compositions during one step of the execution (see for example Figure 2.6, rule $\mathcal{M}_3$).
We have to take care of the fact that function composition has a cost. In order to do so, we assume that step t lasts l time units where l is the length of the longest path in $\mathcal{D}$ joining a node marked at step $t-1$ to a node marked at step t. This is coherent with the assumption that each computation requires 1 time unit. Note that l is also the longest chain of function compositions performed during step t.

Remark 2.3.3. Here is a possible execution of the game. The initial pebbles $\mathcal{A}(0)$ remain unchanged along the execution. An additional pebble is used to mark successively all the nodes in the graph. In this case, marking a node on level n takes $\Omega(n)$ units of time and marking all the nodes up to level n takes $\Omega(n^2)$ units of time. On the other hand, we say that an execution has a *linear* time complexity if it puts a pebble on node (i,n) within $O(n)$ time units for all n. The set of linear executions is not empty. For example, if we mark the nodes as soon as possible, then node (i,n) is marked at time $\lambda n + d_i$.
The executions that we are going to propose to solve Problem 1 will not be optimal in terms of time complexity (i.e. will not be asap). However, they will always be linear, which means that the loss in terms of time efficiency is kept under control (see also the discussion in §2.3.1).

In order for the pebble game to be rigorously defined, we need to have some additional rules. We are going to define four different set of rules $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$.

$\mathcal{M}_1$: Asynchronous execution We consider two additional rules. In particular, we further constrain the rule of Move 1.

- **Move 1'** : Put a pebble on a node if all the predecessors (for the precedence relation) of this node have a pebble.
- It is possible to perform one move of type 1' and several moves of type 2 during one step of the game. On the other hand, it is not possible to perform several moves of type 1'.

Let us consider the example of Figure 2.4. We have represented a small part of the dependence graph of the URE $X_i(n) = F_i(X_1(n-1), X_2(n-1), X_3(n-1)), \forall i = 1, 2, 3$.

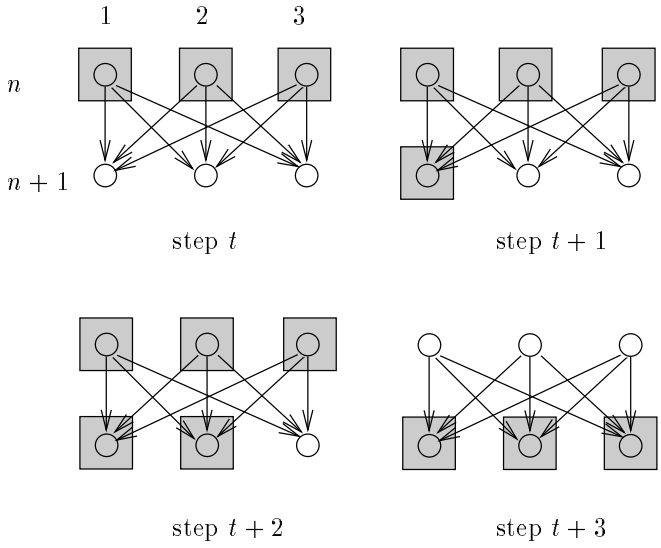


Figure 2.4: Asynchronous rule, $\mathcal{M}_1$, five pebbles are needed.

At step t, there are pebbles on the nodes $(1, n), (2, n)$ and $(3, n)$. At step $t + 1$, we can put a pebble on node $(1, n + 1)$ as all the predecessors (i, n) have a pebble. At step $t + 2$, and for the same reason, we can put a pebble on node $(2, n + 1)$. At step $t + 3$, we put a pebble on node $(3, n + 1)$ and we remove the pebbles on nodes (i, n) (they are not needed anymore). The minimal number of pebbles needed to describe the dependence graph of Figure 2.4 is 5.

This rule corresponds to the necessity of performing asynchronous computations. It is relevant if we use a sequential computer to perform the calculations. In this case, the maximal number of pebbles used during the game corresponds to the minimal number of memory locations needed to carry out the computation.

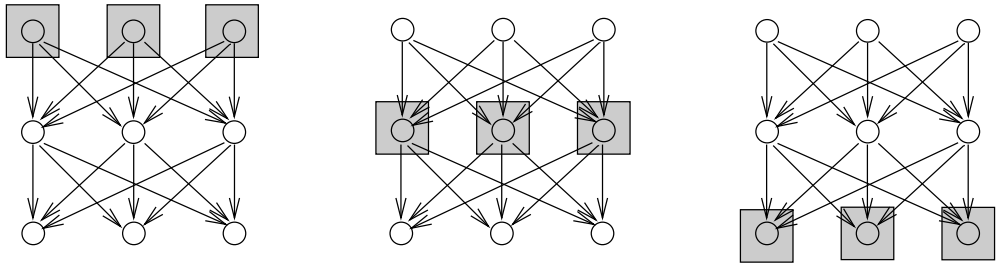
Remark 2.3.4. When Game 1, rule $\mathcal{M}_1$, is performed on a binary tree, the minimal number of pebbles is known as the Strahler's number. This Strahler's number appears in various fields ranging from hydrology or combinatorics to molecular biology. For a nice review paper, the reader is referred to Viennot [135].

In the forthcoming set of rules, we switch back to the original definition of Move 1, see Game 1.

$\mathcal{M}_2$: Synchronous execution We consider Game 1 with the following additional rule.

- Several moves of type 1 and several moves of type 2 can be performed at the same step of the game.

We consider, in Figure 2.5, the same example as previously under the new set of rules.



step t step $t+1$ step $t+2$

Figure 2.5: Synchronous rule, $\mathcal{M}_2$. Three pebbles are needed.

At step n, we have three pebbles on nodes (i,n), $i=1,2,3$. At step $n+1$, we put simultaneously three pebbles on nodes $(i,n+1)$ and we remove the initial pebbles. Hence, the number of pebbles needed by this execution is three.

This rule corresponds to the case where several synchronous processors are used during the computations. It will be adapted if we use parallel synchronous processors with shared memory to carry out the calculations.

$\mathcal{M}_3$: Synchronous regular execution We consider Game 1 with two additional rules

- Several moves of type 1 and several moves of type 2 can be performed at the same step of the game.
- If at step t the configuration is $\mathcal{A}(t)$, then at step $t+1$, the configuration is $(\mathcal{A}(t)+1)$ defined by :

$$(i,n) \in \mathcal{A}(t)+1 \Leftrightarrow (i,n-1) \in \mathcal{A}(t)\ .$$

The example of Figure 2.5 was also verifying the set of rules $\mathcal{M}_3$. To see that $\mathcal{M}_2$ and $\mathcal{M}_3$ are different, let us consider the example of Figure 2.6. It corresponds to the URE $X_1(n) = F_1(X_2(n-1)), X_2(n) = F_2(X_1(n))$.
In Figure 2.6 (I), only one pebble is needed. The corresponding execution verifies rule $\mathcal{M}_2$ but not rule $\mathcal{M}_3$. In Figure 2.6 (II), two pebbles are needed. The corresponding execution verifies rule $\mathcal{M}_3$. The computations are performed according to the following patterns :

- Rule $\mathcal{M}_2$ (Figure 2.6 (I)).
  - step t : $X_2(n) = F_2(X_1(n))$.
  - step $t+1$: $X_1(n+1) = F_1(X_2(n))$.

- step $t+2$: $X_2(n+1) = F_2(X_1(n+1)) \ldots$

- Rule $\mathcal{M}_3$ (Figure 2.6 (II)).
  - step t : $(X_1(n+1), X_2(n+2)) = (F_1 \circ F_2(X_1(n)), F_2 \circ F_1(X_2(n+1)))$.
  - step $t+1$: $(X_1(n+2), X_2(n+3)) = (F_1 \circ F_2(X_1(n+1)), F_2 \circ F_1(X_2(n+2))) \ldots$

Note that in the execution under rule $\mathcal{M}_3$, we perform function compositions, $F_2 \circ F_1$ and $F_1 \circ F_2$. Hence each step lasts two time units.

Rule $\mathcal{M}_3$ has several advantages. First, the number of pebbles needed to carry out the calculations, is easy to compute, it is equal to $|\mathcal{A}(t)|$ (independent of t). Second, a regular execution is interesting for implementation purposes. It provides an easy computation of the variables in the new configuration from the ones in memory by always applying the same operator. A non regular execution could be practically very intricate to implement.

$\mathcal{M}_4$: Synchronous regular non-anticipative execution

- Several moves of type 1 and several moves of type 2 can be performed at the same step of the game.
- If at step t the configuration is $\mathcal{A}(t)$, then at step $t+1$, the configuration is $(\mathcal{A}(t)+1)$ defined by $(i,n) \in \mathcal{A}(t)+1 \Leftrightarrow (i,n-1) \in \mathcal{A}(t)$.
- A path (see Definition 2.4.1) from a node in $\mathcal{A}(t)$ to a node in $\mathcal{A}(t+1)$ contains only nodes belonging either to $\mathcal{A}(t)$ or to $\mathcal{A}(t+1)$.

Let us consider the example of Figure 2.6. In Figure 2.6 (II), we have an example of an execution which verifies rule $\mathcal{M}_3$ but not $\mathcal{M}_4$. For example (see above), the node $(1, n+2)$ is used at step t but is computed only at step $t+1$. On the other hand, in Figure 2.6 (III), we have an execution which verifies rule $\mathcal{M}_4$. In the example of Figure 2.6. The corresponding computation pattern is :

- Rule $\mathcal{M}_4$ (Figure 2.6 (III)).
  - step t : $(X_1(n+1), X_2(n+1)) = (F_1(X_2(n)), F_2 \circ F_1(X_2(n)))$.
  - step $t+1$: $(X_1(n+2), X_2(n+2)) = (F_1(X_2(n+1)), F_2 \circ F_1(X_2(n+1))) \ldots$

Remark 2.3.5. In Figure 2.6, the number of pebbles is the same for the two set of rules $\mathcal{M}_3$ and $\mathcal{M}_4$. It is not always the case, see Figure 2.10.

2.3.3 Summary

In the following, we will use the notations :

- $\mathcal{E}$: the set of all possible (synchronous) executions under rule $\mathcal{M}_2$.

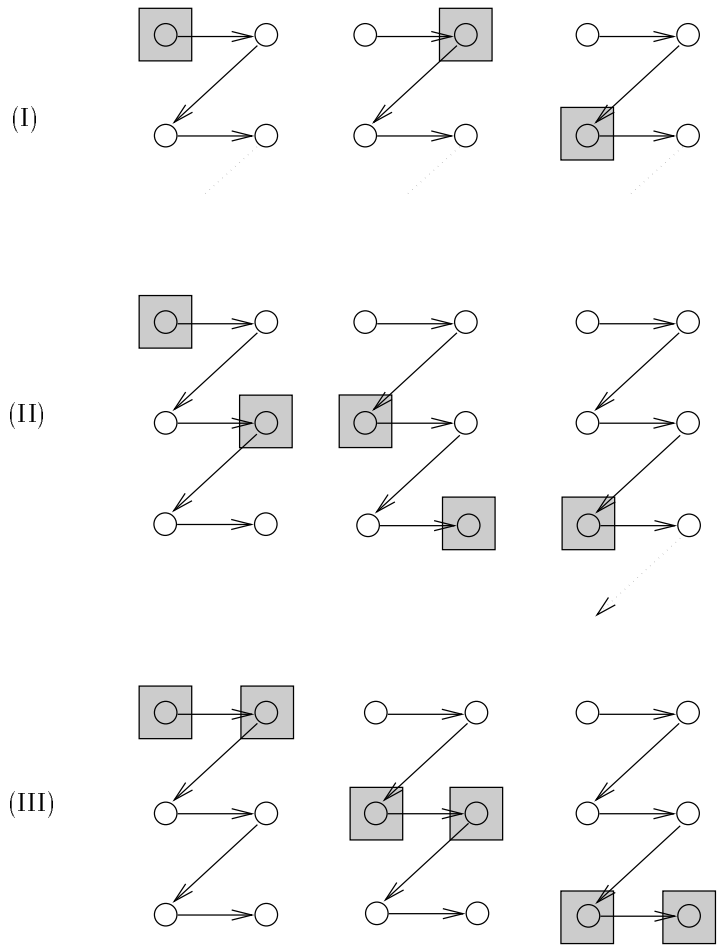


Figure 2.6: Rule $\mathcal{M}_2$ (I), rule $\mathcal{M}_3$ (II) and rule $\mathcal{M}_4$ (III).

- $\mathcal{RE}$: the set of all possible executions under rule $\mathcal{M}_3$. Elements of $\mathcal{RE}$ will be called regular executions.
- $\mathcal{NRE}$: the set of all possible executions under rule $\mathcal{M}_4$. Elements of $\mathcal{NRE}$ will be called non-anticipative regular executions.

Note that

$$\mathcal{NRE} \subset \mathcal{RE} \subset \mathcal{E} .$$

Complexity Results Under rule $\mathcal{M}_1$, the problem of determining the minimal number of pebbles to compute a general directed acyclic graph with one final node has been considered by Sethi [128]. In that paper, it is proved that this problem is NP-complete. Here, we can easily embed any directed acyclic graph on each level of a recycled dependence graph. We also embed the same acyclic graph between two levels of $\mathcal{D}$ (see Figure 2.7).
Now, it is not difficult to see that the minimal number of pebbles needed under rule $\mathcal{M}_1$ in this dependence graph is the minimal number of pebbles necessary to carry out the computation on

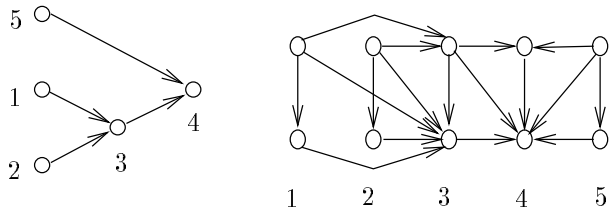


Figure 2.7: Embedding of an arbitrary acyclic graph in a dependence graph.

the original acyclic graph plus the number of columns in $\mathcal{D}$. Therefore, our problem under rule $\mathcal{M}_1$ is also NP-complete.

In the following we will only consider synchronous executions, i.e. the set of rules $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$. In particular, we will characterize executions using a minimal number of pebbles under rules $\mathcal{M}_2$, $\mathcal{M}_3$ and $\mathcal{M}_4$ and we are going to prove that the minimal number of pebbles can be found in polynomial time for the recycled case.

2.4 Cuts and Pebbles

From now on, we consider the recycled case, see §2.2.3. It is always implicitly assumed (unless otherwise specified) that the system under study is recycled.

2.4.1 Cuts in $\mathcal{D}$

Let us recall some classical definitions of graph theory, all defined in the dependence graph, $\mathcal{D}$. For further references, see [65] or [78] for example.

Definition 2.4.1 (path). *A path is a sequence of nodes and arcs in $\mathcal{D}$ of the form $\cdots \to (i_0, n_0) \to (i_1, n_1) \to (i_2, n_2) \to \cdots \to (i_k, n_k) \to \cdots$. A path is bi-infinite if it contains an infinite number of negative nodes and an infinite number of positive nodes.*

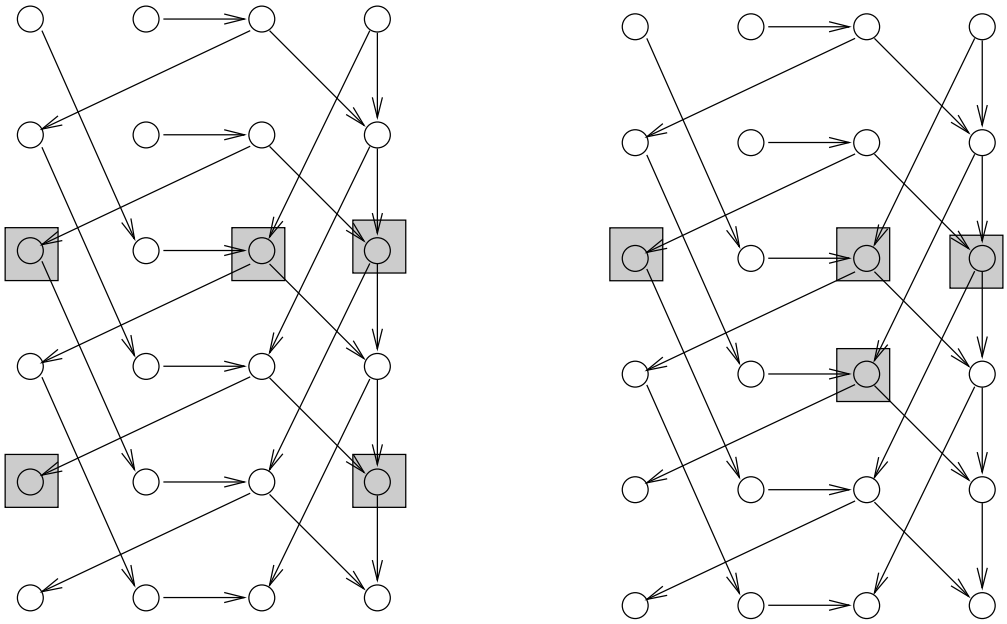
Definition 2.4.2 (cut). *A cut C is a set of nodes in $\mathcal{D}$ such that any bi-infinite path contains at least one node of C. A cut of minimal size is called a minimal cut.*

Definition 2.4.3 (flow). *A flow is a set of bi-infinite paths such that any two paths do not share any node. A flow containing a maximal number of paths is called a maximal flow.*

The most classical notion of cut involves arcs rather than nodes and a flow is a set of paths which do not share arcs rather than nodes. However a small transformation of each node into two nodes connected by an arc would allow us to go back to the original notions.

Definition 2.4.4 (section). *A section S in $\mathcal{D}$ is a set of nodes with exactly one node per column, $S = \{(i, n_i), i \in V\}$.*

Note that since $\mathcal{D}$ is recycled, a cut contains at least one node per column. Using this property, one can define the upper and lower sections of a cut.



A non consecutive cut A consecutive cut

Figure 2.8: Consecutive and non consecutive cuts (on a non-recycled example).

Definition 2.4.5 (upper, lower section). *The upper (resp. lower) section C_u (resp. C_l) of a cut C is the set of nodes (i,n) in C such that the nodes $(i,n-h), h>0$ (resp. $(i,n+h), h>0$) do not belong to C.*

Definition 2.4.6 (consecutive cut). *A cut C in $\mathcal{D}$ is consecutive if on each column of $\mathcal{D}$, C contains only consecutive nodes, i.e.*
For all $i \in V$, $(i,n) \in C$ and $(i,n+1) \notin C \Rightarrow (i,n+k) \notin C, \forall k > 0$.

An example of consecutive and non-consecutive cuts is displayed in Figure 2.8.

Lemma 2.4.7. *There exists a minimal cut of $\mathcal{D}$ which is a minimal consecutive cut.*

Proof. Let C be a minimal consecutive cut. We will prove that C is a minimal cut. First, note that there are no arcs from C_u to $C_l + k, k \geqslant 2$, otherwise C would not be a cut.
Now, consider the sub-graph G of $\mathcal{D}$ made of the nodes $C_u \cup (C_l + 1)$ and the arcs between C_u and $C_l + 1$ in $\mathcal{D}$. See Figure 2.9.
A cut in a finite graph G is a set of nodes such that, when removed from G, there is no arc remaining in G. Let Δ be a cut in G of minimal size. If $|\Delta| < |C_u|$ then $C \backslash C_u \cup \Delta$ would be a consecutive cut in $\mathcal{D}$ strictly smaller than C, which contradicts the fact that C is a minimal consecutive cut. Therefore, we have $|\Delta| = |C_u|$.
An adapted version of a famous "minimax" theorem first proved by Ford-Fulkerson (see [65]) states that we can find $|\Delta|$ node-disjoint arcs in G. Since $|\Delta| = |C_u| = |C_l + 1|$, these arcs define a one to one mapping ϕ from C_u to $C_l + 1$. From ϕ we construct a flow in C in the following

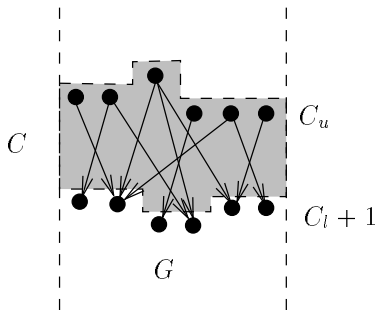


Figure 2.9: Graph G made from the lower and upper sections of C

way. Select all the arcs of the form $((i,n)+k) \to (\phi(i,n)+k)$ for all $(i,n) \in C_u$ and all $k \in \mathbb{Z}$. These arcs form a flow $\mathcal{F}$ in $\mathcal{D}$ of size $|C|$.
Let C_m be a minimal cut in $\mathcal{D}$. Since $\mathcal{F}$ is formed by node-disjoint paths, C_m must contain at least $|\mathcal{F}|$ nodes, $|C_m| \geqslant |\mathcal{F}| = |C|$. We conclude that $|C_m| = |C|$. □

This lemma is interesting by its own. In particular, it gives a proof of the minimax theorem (which exists in many versions) in an infinite graph $\mathcal{D}$.

Corollary 2.4.8. *The size of the minimal cut is equal to the size of the maximal flow in $\mathcal{D}$.*

Another immediate corollary of Lemma 2.4.7 is that there exists a maximal flow $\mathcal{F}$ in $\mathcal{D}$ which is 1-periodic (*i.e.* if the arc $(i,n) \to (j,m)$ belongs to $\mathcal{F}$, then $(i,n+1) \to (j,m+1)$ belongs to $\mathcal{F}$).

Definition 2.4.9. *We say that an arc crosses a section $S = \{(i,n_i), i \in V\}$ downwards if it is an arc of the form $(i,n_i - h) \to (j,n_j + l)$ with $h \geqslant 0$ and $l \geqslant 1$. An arc $(i,n_i + l) \to (j,n_j - h)$ crosses S upwards if $l \geqslant 1$ and $h \geqslant 0$.*

Definition 2.4.10 (compatible section, compatible cut). *A section in $\mathcal{D}$ is compatible if no arc crosses the section upwards. A consecutive cut is said to be compatible if its lower section is compatible.*

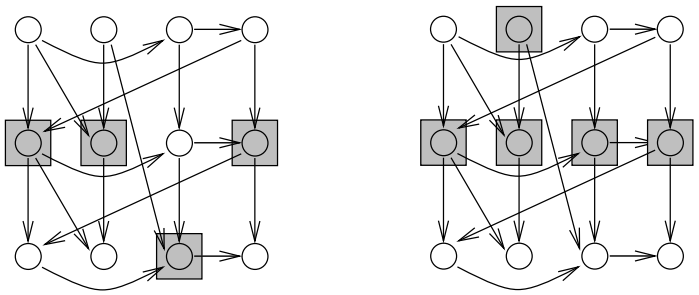
Note that it can be that no minimal consecutive cut in $\mathcal{D}$ is compatible. This is the case in Figure 2.10 where the minimal compatible cut contains 5 nodes while a minimal consecutive cut of size 4 can be found.

2.4.2 Relations with executions of Game 1

Lemma 2.4.11. *A configuration of any execution $e \in \mathcal{E}$ is a cut in $\mathcal{D}$.*

Proof. Let $\mathcal{A}(t)$ be the t-th configuration of some execution e belonging to $\mathcal{E}$. Assume that $\mathcal{A}(t)$ is not a cut. By definition, there exists a bi-infinite path P which does not have any node in $\mathcal{A}(t)$. According to the rule $\mathcal{M}_2$ of Game 1, no positive node on P can ever be marked during the execution after step t and all of them cannot have received a pebble during the t first steps. This contradicts the fact that e is an execution of $\mathcal{E}$. □

Lemma 2.4.12. *A consecutive cut is a configuration of a regular execution ($\mathcal{RE}$).*



Minimal cut Minimal compatible cut

Figure 2.10: Compatible and non compatible cuts.

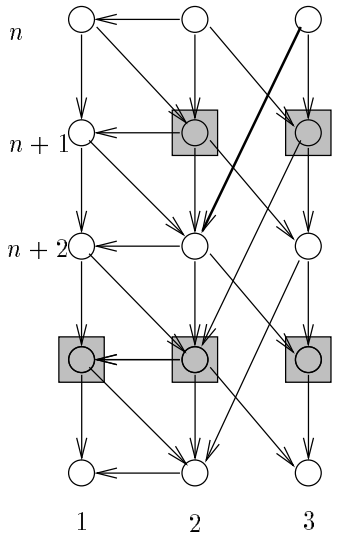


Figure 2.11: A non consecutive cut which is not a regular configuration.

Proof. Let C be a consecutive cut in $\mathcal{D}$. We have $C+1 = (C \backslash C_u) \cup (C_l + 1)$. Therefore, C is a regular configuration if and only if for each node (i,n) in $C_l + 1$, there is no infinite path P terminating in (i,n) that does not intersect the cut. But if such a path could be found, then the bi-infinite path $P \cup \{(i, n+h), h \in \mathbb{N}\}$ would not intersect C. This contradicts the fact that C is a cut. □

Remark 2.4.13. Note that a non consecutive cut may not be a regular configuration, as illustrated in the example of Figure 2.11. In that example, the node $(2, n+2)$ belongs to $C+1$ but cannot be computed using only variables in C (as it depends on $(3,n)$ for example). Therefore, the cut C is not a regular configuration.

Lemma 2.4.14. *A compatible cut is a configuration of a non-anticipative regular execution ($\mathcal{NRE}$).*

Proof. Let C be a compatible cut. Since C is consecutive by definition, Lemma 2.4.12 tells us that C is a configuration $\mathcal{A}(t)$ of a regular execution e. Suppose that e is anticipative. This means that there exists a node (say (i,n)) in $\mathcal{A}(t+1)$ with a predecessor (say (j,m)) not marked at steps t or $t+1$. The arc $(j,m-1) \to (i,n-1)$ crosses the lower section of C upwards. This contradicts the fact that C is compatible. □

We are now ready to state the main result of this section.

Theorem 2.4.15. *Let us consider a recycled system of URE $\mathcal{S}$. We perform Game 1 on its associated dependence graph $\mathcal{D}$. We have*

$$\min_{e \in \mathcal{RE}} \mathcal{P}(e) = \min_{e \in \mathcal{E}} \mathcal{P}(e). \tag{2.5}$$

In other words, there exists a regular execution which requires a minimal number of pebbles.

Proof. The proof of the theorem is a direct consequence of Lemmas 2.4.11, 2.4.12 and 2.4.7. First, note that all configurations are cuts, Lemma 2.4.11. Let C be a consecutive cut of minimal size, which exists by Lemma 2.4.7. By Lemma 2.4.12, C is a regular configuration. □

Theorem 2.4.15 has several interesting corollaries. First, it allows one to focus on regular executions since no fancy irregular execution of the URE can be done with fewer pebbles. Then, it provides a polynomial method to find an optimal execution as shown in §2.4.3. As for non-anticipative executions, polynomial algorithms will be given in §2.6.4.

2.4.3 Complexity results

We are going to compute a maximal flow in $\mathcal{D}$ and then apply Corollary 2.4.8. If we want to use the algorithm of Ford and Fulkerson [65] to compute a maximal flow in $\mathcal{D}$, we need first to restrict ourselves to a finite graph.

We call *span* of a cut the difference between the smallest level and the largest level containing a node of the cut.

A *slice* of $\mathcal{D}$ from level 0 to level n will be sufficient to compute the maximal flow in the graph if a consecutive minimal cut spans over less than n levels. So it is important to determine, or at least to bound, the span of a consecutive minimal cut.

Lemma 2.4.16. *The span of a minimal consecutive cut is smaller than the total sum of the delays in $\mathcal{R}$.*

Proof. Let C be a minimal consecutive cut. The associated maximal 1-periodic flow $\mathcal{F}$ is a set of paths in $\mathcal{D}$. First, note that these paths cover all the nodes in $\mathcal{D}$. Indeed, by the 1-periodicity of $\mathcal{F}$, if a node (i,n) is not in $\mathcal{F}$, then the whole column $(i,.)$ is not in $\mathcal{F}$, but this means that the bi-infinite path $\{(i,n), n \in \mathbb{Z}\}$ can be added to flow $\mathcal{F}$ and this would contradict the maximality of $\mathcal{F}$.
Let P_1 be any path in $\mathcal{F}$. We deduce from the 1-periodicity of $\mathcal{F}$ that P_1 is periodic. Let $i_0, i_1, \cdots, i_{l_1}, i_0, i_1, \ldots$ be the successive columns visited by the path P_1. Let (i_0,n) and $(i_0, n+k_1)$ be the consecutive nodes visited by the path P_1 on column $(i_0,.)$. Using the 1-periodicity of $\mathcal{F}$, the total number of paths intersecting columns $i_0, i_1, \cdots, i_{l_1}$ in $\mathcal{F}$ is k_1. It implies that the

span of C on columns $i_0, i_1, \cdots, i_{l_1}$ is smaller than k_1. By definition of $\mathcal{R}$, there exists a circuit (L_1) in $\mathcal{R}$ containing the nodes $i_0, i_1, \cdots, i_{l_1}$ and of total delay k_1. The span of C on columns $i_0, i_1, \cdots, i_{l_1}$ is smaller than k_1.

A new path P_2 in $\mathcal{F}$ not intersecting columns $i_0, i_1, \cdots, i_{l_1}$ ranges over different columns

$$i_{l_1+1}, i_{l_1+2}, \cdots, \cdots, i_{l_2},$$

and defines a circuit (L_2) in $\mathcal{R}$ similarly. The span of C on columns $i_{l_1+1}, \cdots, i_{l_2}$ is smaller than k_2. We apply the same arguments to all the paths in $\mathcal{F}$ until all columns in $\mathcal{D}$ are covered. This defines a set of circuits H in $\mathcal{R}$ covering $\mathcal{R}$.

We build a new graph $\mathcal{G}$ starting with $\mathcal{R}$ and where each circuit in H is aggregated into one node. The graph $\mathcal{G}$ has $|H|$ nodes and the arcs of $\mathcal{G}$ do not belong to any circuit in H.
The span of C is smaller than the sum of the spans on all the circuits in H plus the sum of the delays on all the arcs in $\mathcal{G}$. Note that no delay is counted twice in this upper bound. Therefore, the total span (M) of C is smaller than the total sum of the delays in $\mathcal{R}$. □

Remark 2.4.17. This bound is tight since it is not difficult to exhibit examples in which the span of the minimal cut is the sum of all the delays in $\mathcal{R}$. However, in most cases, the span of a minimal consecutive cut is significantly smaller.

Let M be the sum of the delays in $\mathcal{R}$. A slice of $\mathcal{D}$ with M levels has the same cut size as $\mathcal{D}$ itself. The computation of the minimal cut in a finite slice of $\mathcal{D}$ can be done using the augmenting path algorithm, see [65] [78]. Starting with a 1-periodic flow (the recycled columns) and maintaining the 1-periodicity throughout the construction yields a maximal 1-periodic flow. The complexity of the construction of the maximal flow is $O(M^2k^2)$. By Corollary 2.4.8, it provides the size of a minimal (consecutive) cut in $\mathcal{D}$.

2.5 Cuts and Delays

In this section, we will exhibit the relations that can be found between cuts in $\mathcal{D}$ and values of the delays in $\mathcal{R}$.

2.5.1 Retiming

A *retiming* of $\mathcal{R}$ is a transformation of the graph resulting in a decrease or increase of the values of the delays but with no transformation of the graph topology. This notion has been described in digital circuits to move registers (see §2.6) and in Petri nets, where a retiming corresponds to the firing of transitions (see §2.7).

Definition 2.5.1 (retiming). *A retiming of $\mathcal{R}$ is a node function $r : V \to \mathbb{N}$ which specifies a new value of the delays. After retiming r, the value of the delay on an arc (i, j) in the new graph $\mathcal{R}_r$ is $\gamma_r = \gamma + r(i) - r(j)$.*

In the example of Figure 2.12, the new values of the delays correspond to a retiming r such that $r(1) = 1, r(2) = 1$ and $r(3) = 0$.
Note that after a retiming r of $\mathcal{R}$ the delay on one arc can be negative as in Figure 2.12.

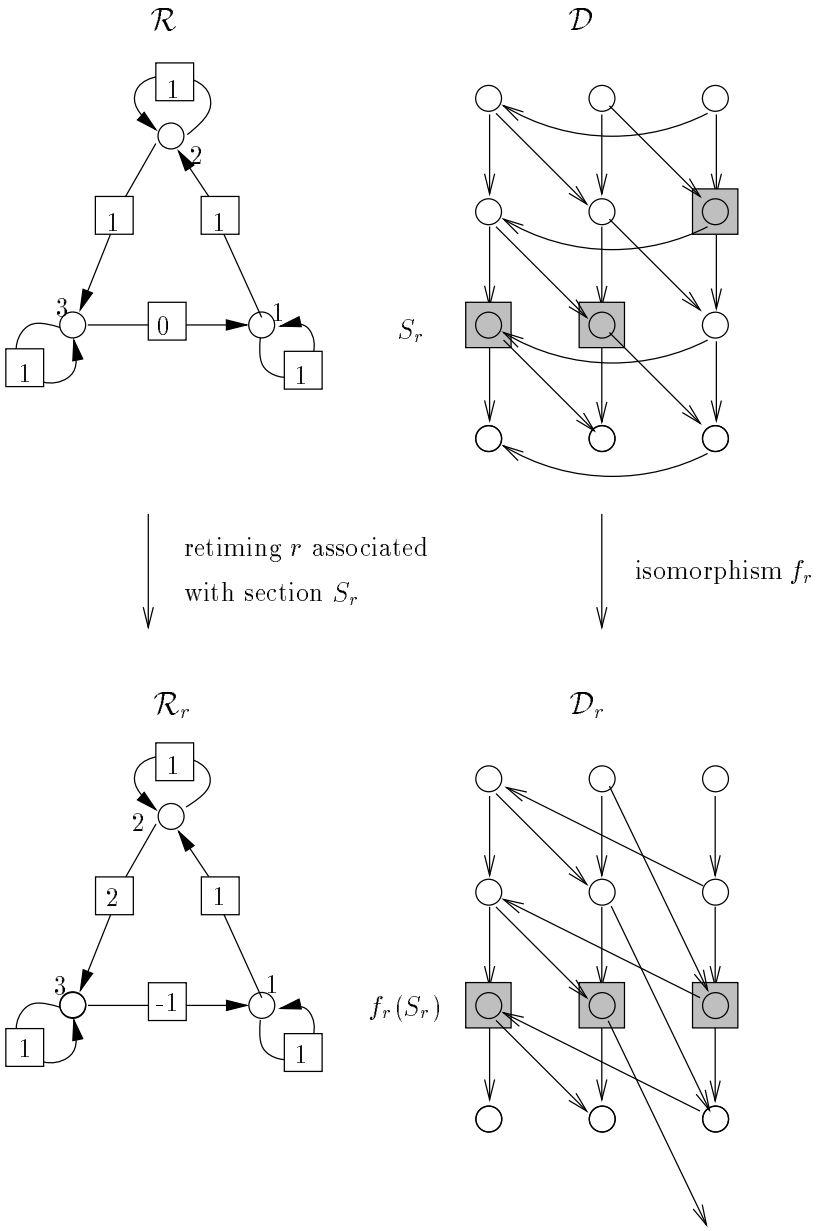


Figure 2.12: Retimed reduced graph and dependence graph.

Lemma 2.5.2. *Two retimings r and r' yield the same value of the delays in a connected graph $\mathcal{R}$ if and only if there exists a constant $h \in \mathbb{Z}$ such that $\forall i \in V, r(i) = r'(i) + h$.*

Proof. First, if $r(i) = r'(i) + h$ for all $i \in V$, then on any arc (i,j), $\gamma_r = \gamma + r(i) - r(j) = \gamma + r'(i) - r'(j) = \gamma_{r'}$. Conversely, if $\gamma_{r'} = \gamma_r$, then $r(i) = r'(i) + h$ and $r(j) = r'(j) + h$ for some $h \in \mathbb{Z}$. The fact that $\mathcal{R}$ is connected implies that the constant h is the same for all the nodes in V. □

The question that arises now is what is the corresponding notion in the graph $\mathcal{D}\Gamma$ To answer this question, let us consider the graph $\mathcal{D}_r$ associated with the retimed reduced graph $\mathcal{R}_r$. This dependence graph can be constructed directly from $\mathcal{D}$ by shifting the columns as described in Lemma 2.5.3.

Lemma 2.5.3. *A retiming r in $\mathcal{R}$ corresponds to an isomorphism f_r between $\mathcal{D}$ and $\mathcal{D}_r$ defined by:*

$$\begin{array}{rccc} f_r : & \mathcal{D} & \rightarrow & \mathcal{D}_r \\ & (i,n) & \rightarrow & (i, n - r(i)) \end{array}$$

The function f_r will also be called a retiming of $\mathcal{D}$.

Proof. By definition of $\mathcal{D}_r$, there is an arc from (i,n) to (j,m) in $\mathcal{D}_r$ if the delay in $\mathcal{R}_r$ on arc (i,j) is $\gamma_r = m - n$. We have $\gamma_r = \gamma + r(i) - r(j) = (m - r(j)) - (n - r(i))$. Therefore, f_r is an isomorphism between $\mathcal{D}$ and $\mathcal{D}_r$. □

We recall that the notion of *section* was defined in 2.4.4. A retiming r in $\mathcal{R}$ can be associated with the section $S_r = \{(i, r(i)), i \in V\}$ in $\mathcal{D}$.
Lemmas 2.5.2 and 2.5.3 tell us that two retimings r and r' are similar (in the sense that they yield the same value of the delays) if and only if they are associated with two sections S_r and S'_r with $S_r = S_{r'} + h$, for some $h \in \mathbb{Z}$. This relation enables us to define a parallelism relation between sections in $\mathcal{D}$ as well as between retimings in $\mathcal{R}$. We say that section S_r (resp. retiming r) is equivalent to section $S_{r'}$ (resp. retiming r') if $S_r = S_{r'} + h$, for some $h \in \mathbb{Z}$. In the following, we will always consider one arbitrary section among the equivalence class and call it the section associated with retiming r.

2.5.2 Counting the delays

Given a graph $\mathcal{R}$, there are different possible ways to count the number of delays involved in the graph. We are going to propose two different ways of counting, mode A and mode B.

Mode A : The number of delays in $\mathcal{R}$ is

$$\Gamma_A = \sum_{i \in V} \sum_{(j,\gamma) \in \mathcal{D}_i} \gamma \,. \tag{2.6}$$

Mode A corresponds to the exact number of registers appearing in the graphical representation of the reduced graph $\mathcal{R}$ as defined in §2.2.2. See for example, Figure 2.13 (A).

Mode B : The number of delays in $\mathcal{R}$ is

$$\Gamma_B = \sum_{j \in V} \max\{\gamma \mid \exists i \text{ s.t. } (j,\gamma) \in \mathcal{D}_i\} . \tag{2.7}$$

Let us explain this mode of counting. First, remark that $\Gamma_B \leqslant \Gamma_A$. Assume that node $j \in V$ has several output arcs with respective delays $\gamma_1, \gamma_2, \cdots, \gamma_l$. If we allow the possibility to share the delays between these l arcs, the number of delays will be counted as $\max_k \gamma_k$ instead of $\sum_k \gamma_k$ as in Mode A.

Graphically, Γ_B corresponds to the number of delays in a modified reduced graph where we have performed a forward splitting of the nodes. An example is provided in Figure 2.13 (B). We have added a "dummy" node, represented by a black dot, with function F =Identity. The other nodes remain unchanged. This reduced graph describes exactly the same system of URE. More precisely, the variables computed at the white nodes in Figure 2.13 (B) are the same as the variables computed in Figure 2.13 (A).

There is another way to interpret mode B. Let us assume for a moment that we modify the definition of a reduced graph. We consider a reduced graph $\tilde{\mathcal{R}}$ where delays are put on nodes instead of arcs. The total number of delays in graph $\tilde{\mathcal{R}}$ is equal to Γ_B. This is illustrated in Figure 2.13 $(\tilde{B})$.
In the following, we will say, with some abuse of language, that counting mode A corresponds to delays on arcs and counting mode B to delays on nodes.

Other ways of enumerating delays are conceivable. We will not consider them as they appear to be less interesting, mathematically speaking as well as from a practical point of view.

Deciding which counting mode of the delays to choose is very important. Different modes will yield different optimal graphs, after minimization of the number of delays.

2.5.3 Cuts and delays

We recall that the system under study is assumed to be recycled.

Consider a section $S = \{(i, n_i), i \in E\}$ in $\mathcal{D}$. We define a consecutive cut $\mathcal{C}(S)$ of the graph in the following way. We define the set $\mathcal{C}(S)$ in the following way :

$$\mathcal{C}(S) \stackrel{\text{def}}{=} \{(i,n), i \in V, n \leqslant n_i \mid \exists j \in V, m > n_j, (i,n) \to (j,m)\} .$$

Note that $\mathcal{C}(S)$ is a cut with lower section S. Furthermore, if any node is removed from the upper section of $\mathcal{C}(S)$, then it is not a cut anymore.

In a cut C, a node $(i,n) \in C$ is *redundant* if $C \backslash \{(i,n)\}$ is a cut. Note that any consecutive cut C with no redundant node on its upper section is characterized by its lower section S_l only. More precisely, we have $C = \mathcal{C}(S_l)$.
We are now ready to state the relations between delays in $\mathcal{R}$ and consecutive cuts in $\mathcal{D}$.

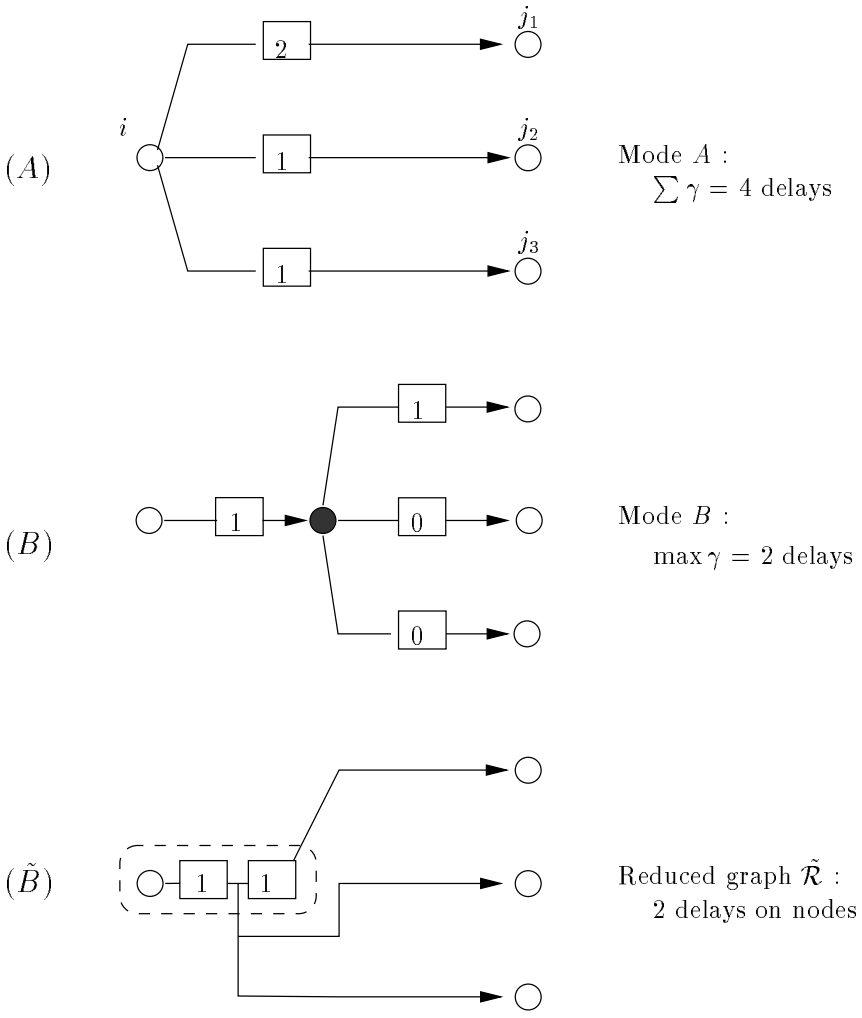


Figure 2.13: Different ways to enumerate the delays.

Lemma 2.5.4. *Let r be a retiming of $\mathcal{R}$ and S_r an associated section in $\mathcal{D}$. Then the number of delays in $\mathcal{R}_r$ under mode B (Γ_B) is equal to the cardinal of the cut $\mathcal{C}(S_r)$.*

Proof. We recall that the section associated with r is $S_r = \{(i, r(i)), i \in V\}$. First, let us prove that the size of $\mathcal{C}(S_r)$ (cut in $\mathcal{D}$) is the size of $\mathcal{C}(f_r(S_r))$ (in $\mathcal{D}_r$). Note that by definition, $f_r(S_r)$ is on a single level, $f_r(S_r) = \{(i, 0), i \in V\}$. From the definition of f_r, it should also be clear that $f_r(\mathcal{C}(S_r)) = \mathcal{C}(f_r(S_r))$ and *a fortiori* $f_r(\mathcal{C}(S_r))$ and $\mathcal{C}(f_r(S_r))$ have the same size.
It remains to be shown that the size of $\mathcal{C}(f_r(S_r))$ is the number of delays counted according to mode B. The delay on node i of $\mathcal{R}$ is the maximum of all γ_r for all (j, γ_r) in $\mathcal{D}_{i,r}$. This maximum (m) induces an arc in $\mathcal{D}_r$ from node $(i, -m+1)$ to node $(j, 1)$. By construction, $\mathcal{C}(f_r(S_r))$ contains exactly m nodes on column i: nodes $(i, 0), (i, -1), \cdots, (i, -m+1)$.
The same argument repeated on each column of $\mathcal{D}_r$ finishes the proof. □

We recall that given a section S, we defined the arcs crossing S upwards or downwards in Definition 2.4.9.

Lemma 2.5.5. *Let r be a retiming of $\mathcal{R}$ and S_r an associated section in $\mathcal{D}$. The number of delays in $\mathcal{R}_r$ under mode A (Γ_A) is equal to the number of arcs in $\mathcal{D}$ crossing section S_r downwards minus the number of arcs crossing S_r upwards.*

Proof. By definition of f_r, all the arcs crossing $f_r(S_r)$ in $\mathcal{D}_r$ are the transform by f_r of the arcs crossing S_r in $\mathcal{D}$. We will rather count the arcs in $\mathcal{D}_r$. Pick one arc in $\mathcal{R}_r$ (i,j) with delay $\gamma \geqslant 0$. In $\mathcal{D}_r$, this arc induces exactly γ arcs crossing $f_r(S_r)$ downwards, the arcs:

$$\begin{array}{rcl} (i,0) & \to & (j,\gamma) \\ (i,-1) & \to & (j,\gamma-1) \\ & \vdots & \\ (i,-\gamma+1) & \to & (j,1). \end{array}$$

Similarly, an arc in $\mathcal{R}_r$, (i,j) with delay $\gamma < 0$ induces exactly $-\gamma$ arcs crossing $f_r(S_r)$ upwards:

$$\begin{array}{rcl} (i,-\gamma) & \to & (j,0) \\ (i,-\gamma-1) & \to & (j,-1) \\ & \vdots & \\ (i,1) & \to & (j,\gamma+1). \end{array}$$

The same argument applied to all the columns finishes the proof. □

Remark 2.5.6. Lemma 2.5.4 shows that one can describe the number of delays Γ_B as the cardinal of a set of nodes of $\mathcal{D}$. On the other hand, we deduce from Lemma 2.5.5 that the number of delays Γ_A is computed as the cardinal of a set of arcs in $\mathcal{D}$. This is natural as we have seen that (roughly speaking) mode A corresponds to delays on arcs and mode B to delays on nodes.

The notion of compatible cut introduced in Definition 2.4.10 has a very natural interpretation in terms of delays.

Proposition 2.5.7. *Let r be a retiming of $\mathcal{R}$ and S_r an associated section in $\mathcal{D}$. The retimed reduced graph $\mathcal{R}_r$ has only non-negative delays if and only if the section S_r is compatible in $\mathcal{D}$.*

In Figure 2.14 (this example is the same as the one of Figure 2.10), we have represented the retimed reduced graphs associated with two sections (cuts). One of them is compatible, Figure 2.14 (I), and the other one is not compatible, Figure 2.14 (II).
In §2.4.2, we have established the relations between configurations (for Game 1) and cuts in the dependence graph. In this paragraph, we have established the relations between cuts and delays. As an immediate by-product, we obtain the relations between delays and pebble configurations.

- An execution $e \in \mathcal{E}, e \notin \mathcal{RE}$ has configurations with different shapes at each step. Since any configuration can be viewed as a set of value of the delays in a retimed reduced graph, then an execution which is not regular provides a different value of the delays in $\mathcal{R}$ at each step of the computation.
- On the contrary, for an execution $e \in \mathcal{RE}$ the configurations are just shifted between two steps and this induces a fixed value of the delays in $\mathcal{R}$.
- Finally, an execution $e \in \mathcal{NRE}$ corresponds to fixed and non-negative delays.

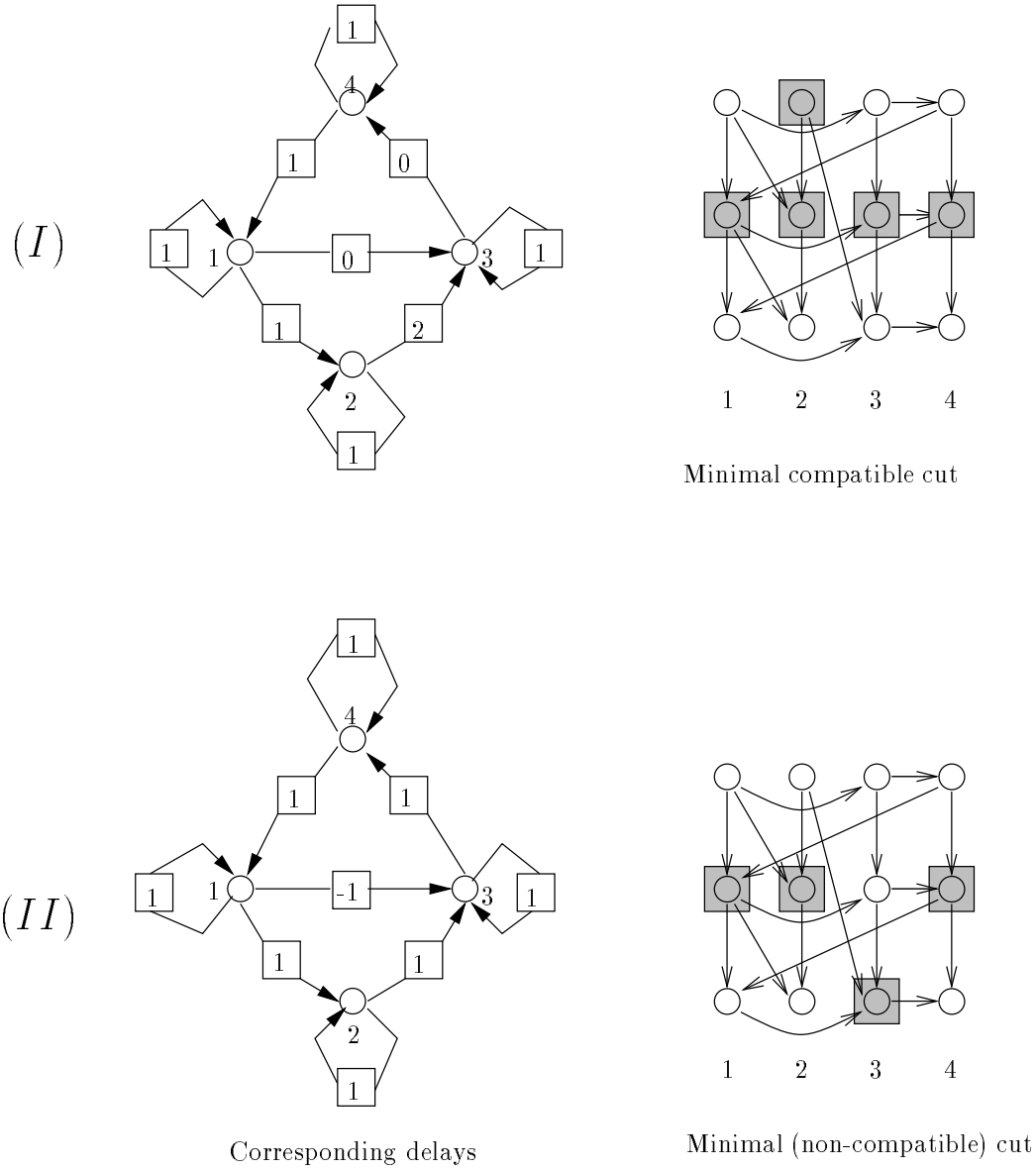


Figure 2.14: Compatible and non compatible cuts, non-negative and negative delays.

2.5.4 Summary and open problems

In the recycled case, the following table gives a summary of the main relations established so far between executions of a system of URE, cuts in $\mathcal{D}$ and delays in $\mathcal{R}$.

Executions of Game 1	Cuts in $\mathcal{D}$	Delays in $\mathcal{R}$
execution in $\mathcal{E}$	arbitrary cut	changing delays
regular execution, $\mathcal{RE}$	consecutive cut	fixed delays
non-anticipative reg. exec., $\mathcal{NRE}$	compatible cut	non-negative fixed delays

To complete the picture, it would be nice to extend all the results presented in this section to the non recycled case. The different definitions (cut, flow, section) extend easily to the non-recycled case. The main results which would make everything else easy to generalize are of two types. Results related with cuts in $\mathcal{D}$, §2.4.1, and results linking cuts and regular configurations, §2.4.2. For example, is it possible to find a minimal cut which is consecutive (generalization of Lemma 2.4.7) Γ Can we find a minimal consecutive cut which is a regular configuration (generalization of Lemma 2.4.12)Γ It seems that most of these properties still hold in the non-recycled case but at this point the problem is still open.

One of the main results so far is that the size of the minimal consecutive cut is the same as the size of the minimal cut in $\mathcal{D}$ (see Lemma 2.4.7). However, the example displayed in Figure 2.14 shows that in some cases this minimal consecutive cut is not compatible and therefore, its associated execution is anticipative. Although anticipative executions seems to have no or little interest in practice, we will show in the two applications presented below (sections §2.6 and §2.7) that there are particular situations in which they can be used efficiently.

2.6 Application 1 : Registers in Circuit Design

In this section we will show how the previous results relate to the problem of register minimization in digital circuits. The interest of the relation will be two-fold. In a first part we show how the notions we defined so far help to prove the optimality of retiming in digital circuits. In a second part, we will use the algorithms developed in digital circuits to get optimal regular executions of a system of URE.

2.6.1 Definition of a circuit

A digital circuit is constituted by functional gates, wires and registers. More precisely,

- A functional element computes an output data from one or several input data. For example, in the case of a logical circuit, the functional elements will be boolean logical gates (AND, OR,...)
- A wire between element i and element j enables to transfer the output data of i which becomes an input data for j.
- A register corresponds to a storage facility. A register of size p (or equivalently p registers) between elements i and j enables to keep in memory the last p values computed by the element i.

The model of the behavior of the system is the following. There is a global clock for the system. Between two clock ticks, here are the operations taking place.

- Functional element :
  1. receive the input data from upstream registers.
  2. compute a new output data.
  3. send the output data to downstream registers.
- Register :
  1. transmit the stored data downstream (to another register or a functional element depending on the structure).
  2. remove the stored data.
  3. receive a new data from upstream (from another register or a functional element depending on the structure).

Between two clock ticks, these operations are synchronously performed at *all* functional elements and registers[1].

Let $X_i(n)$ be the n-th variable computed at element i. After n clock ticks, exactly n values have been computed at each element i, i.e. the variables $\{X_i(m), m \leqslant n\}$ have been computed. The number of registers on a wire between i and j corresponds to the number of variables $X_i(n-k)$ which need to be still in the memory in order to carry on the computation of the variables $X_j(n+m), m \in \mathbb{N}$.
It appears from the previous description that a digital circuit can be viewed as the reduced graph $\mathcal{R}$ of some system of URE. The functional elements of the circuit correspond to the nodes of $\mathcal{R}$, the wires to the arcs and the registers to the delays. The computation operation corresponding to the functional element i is denoted by F_i to be coherent with previous notations. In the remainder of the section, we will use indifferently the terminology of digital circuits and the one of reduced graphs.

We have represented an example of a digital circuit in Figure 2.15. We have represented the flow of data between clock ticks. We have chosen on purpose a graphical convention coherent with the one of reduced graphs.

Remark 2.6.1. It might be interesting to consider that the different operations described above have a duration, let us say 1 unit of time for a computation and instantaneous for a transmission-reception (same assumption as in Remark 2.3.3). According to this, the elementary operations occurring between two ticks have a total execution time. It is 1 for the system $X_j(n) = X_i(n-\gamma), \gamma > 0$ and 2 for the system $X_j(n) = X_i(n)$ (both variables are computed successively during the same clock interval). Hence the length of the time interval between two clock ticks has to be at least 1 in the first case and at least 2 in the second one. More generally the time interval between two clock ticks has to be equal (at least) to the length (i.e. the number of nodes) of the longest path without registers in the graph $\mathcal{R}$. Hence it is often a problem of

[1] In particular, we do not consider systems where the computations times might be different from one element to the other.

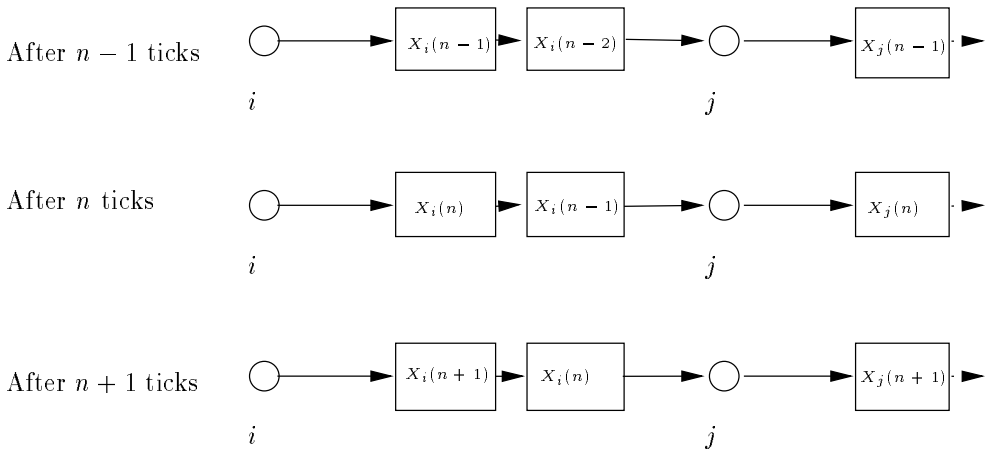


Figure 2.15: Digital circuit computing $X_j(n) = F_j(X_i(n-2), \ldots)$.

practical interest to minimize the longest path without registers, see [100] [34] [73]. We will not consider this problem in the following. We consider the problem called min-area in [100]. It consists in minimizing the number of registers.

2.6.2 Counting the registers

The two modes for counting delays (see §2.5.2) are interesting from a practical point of view, when delays are viewed as registers in digital circuits. In order to explain it, we are going to focus on the example of Figure 2.13. Let us compare the characteristics of the three digital circuits, A, B and $\tilde{B}$, proposed in Figure 2.13. In circuit (A), there is a synchronous *write* operation (also called fanout) performed by node i when displaying its output data to downstream registers j_1, j_2 and j_3. In circuit $(\tilde{B})$, there is a synchronous *read* operation performed by the nodes j_2 and j_3 when they get the variable stored in the first register of node i. According to physical and technological constraints, it might be better to avoid either synchronous read or synchronous write, hence to prefer either circuit (A) or $(\tilde{B})$.
Even if we assume that synchronous read has to be avoided (as it is often the case for digital circuits), we might be interested in considering circuit (B) instead of circuit (A). In circuit (B), we have less registers but more functional gates. Hence depending on the compared cost of a node and a register, one shall consider one circuit or the other.

2.6.3 Minimizing the registers

A classical problem in circuit design is to minimize the number of registers used while preserving the functional behavior of the circuit (i.e. while computing the same variables $X_i(n)$), see the seminal paper of Leiserson and Saxe [100]. If the circuit (i.e. the corresponding reduced graph) is recycled, this problem is directly connected with the notions introduced in §2.4 and §2.5. It enables us to propose some complements to the results of [100] for the special case of recycled circuits.

Optimality of retiming

In [100], Leiserson and Saxe define a notion of retiming which is exactly the one of Definition 2.5.1. They restrict their attention to *legal* retimings.a

Definition 2.6.2. *A retiming r is legal if $\mathcal{R}_r$ has only non-negative delays.*

This is a natural restriction as legal retimings are the only one having a physical meaning for circuits (at least apparently, see §2.6.3). They also define register sharing. This corresponds exactly to the transformation from circuit (A) to circuit (B) in Figure 2.13. Leiserson and Saxe prove that retiming and register sharing preserve the functional behavior of the circuit. Then they propose an algorithm to compute the optimal circuit after retiming and also after retiming and register sharing, see §2.6.4.

However the question whether other techniques can be used to get a circuit with even fewer registers remains to be answered. Using the results of previous sections, we show that the retiming technique combined with register sharing does in fact minimize the number of registers.

This result comes from the following argument. Let us consider a circuit. We consider the same circuit where we have positioned the registers in an arbitrary way. We assume that the functional behavior is not modified. These registers can be seen as delays in the reduced graph $\mathcal{R}$. If the functional behavior of the circuit is preserved, it means that the delays correspond to a non-anticipative regular configuration in the associated dependence graph $\mathcal{D}$. But in the recycled case, such a configuration is also a compatible cut in $\mathcal{R}$. The lower section of this cut defines a compatible section which is in turn associated with a legal retiming of the circuit. Therefore, the positions of the registers can be obtained from a retiming of the original circuit.

Corollary 2.6.3. *If we count the number of registers according to mode B, then we can obtain a circuit with a minimal number of registers solely by performing retiming.*

Proof. This is a direct consequence of Lemma 2.5.4. □

Further modifications of the circuit

If we allow other modifications of the circuit than just register sharing, further improvements on the number of registers can be obtained.

Let us consider the best possible retiming in the original circuit without restricting ourselves to legal retimings (Definition 2.6.2). It corresponds to a minimal consecutive (but not necessarily compatible) cut in the associated dependence graph $\mathcal{D}$, see §2.5.4.
It is possible to perform some appropriate modifications to the structure of the circuit to go back to positive delays. The procedure is as follows. For each node i following a negative delay, we track back the paths terminating at node i until the total delay on each path is non-negative. This is always possible for circuits associated with constructive URE. The nodes initiating such paths are duplicated into 2 nodes computing the same function. In the example of Figure 2.16, we have to track back two paths : $3 \leftarrow 1 \leftarrow 1$ and $3 \leftarrow 1 \leftarrow 4$ and node 1 is duplicated into 1 and $1'$.
If the registers are counted according to mode B, the resulting circuit uses only 4 registers while the best possible number of registers we can get with only legal retimings is 5. Of course, on the other hand, we have to increase the number of functional nodes.

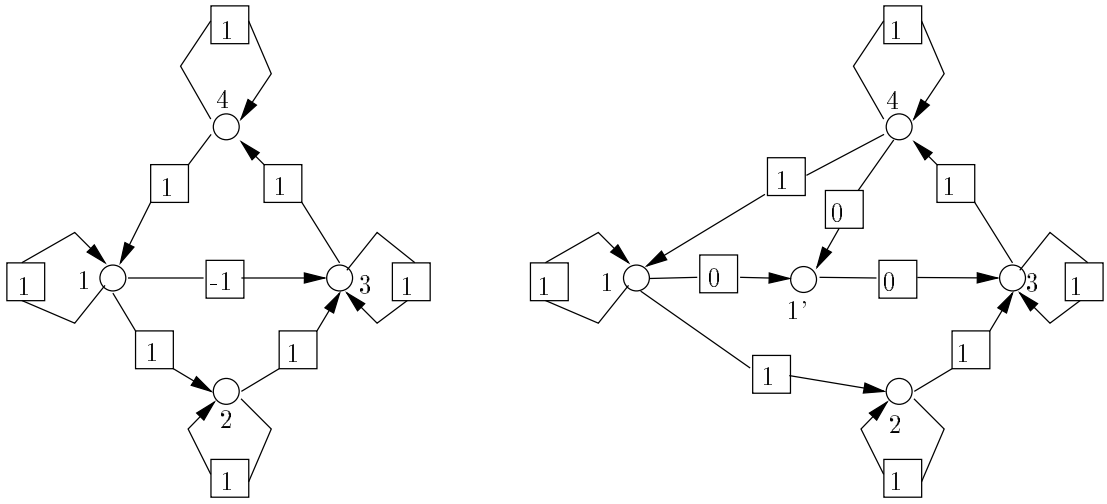


Figure 2.16: A transformation of a circuit with negative delay into a circuit with non-negative delay.

2.6.4 Complexity results

In [100], an algorithm is given to compute the best legal retiming of a circuit (with or without using the register sharing technique). The complexity of this algorithm is $O(|E|^2 log|V|)$. An efficient implementation of this algorithm can be found in [129]. In the recycled case, this algorithm can be used to compute the minimal compatible cut in $\mathcal{D}$, using corollary 2.6.3. And finally using the correspondence between compatible consecutive cuts and non-anticipative regular executions, see Lemma 2.4.14, this also gives a way to find an optimal non-anticipative regular execution of a system of URE. This is an example of results originally proved for digital circuits and applied in the context of URE.
Further results developed for circuits can be applied in the computation of URE. It is the case of the problem of the minimization of the clock period in digital circuits, see Remark 2.6.1. This issue is not addressed here.

Conversely, the results of §2.4.3 (using the Ford-Fulkerson algorithm) can be applied in the context of digital circuit to compute the optimal (non necessarily legal) retiming. This is interesting as it is not straightforward to extend the original algorithm of Leiserson and Saxe to general retimings.

2.7 Application 2: (max,+) Linear Systems and Parallel Simulation

In the following we will apply our results to a particular class of URE: $(\max, +)$ linear systems, and to issues arising in the distributed simulation of such systems.
Our interest for $(\max, +)$ systems comes originally from the analysis of a class of timed Petri nets: Timed Event Graphs [8]. However, these systems arise naturally in the study of general URE.

To see this, assume that, in the computation of some URE of the form (2.2), computing $X_i(n) = F_i(X_j(n-\gamma),\ldots)$ requires σ_i units of time. Assume moreover that the computation is performed on a parallel computer with an unlimited supply of identical processors, common memory (or instantaneous communication), and no synchronization overhead. If variables are computed as soon as possible (greedy execution), then the *makespan* of the computation is given by an URE with the same structure as (2.2). Indeed, if $T_i(n)$ is the instant at which the computation of $X_i(n)$ starts, then T_i is given by $T_i(n) = \max(T_j(n-\gamma)+\sigma_j,\ldots)$.
We shall discuss below some issues arising in the computation of $(\max,+)$ systems. It should be clear that the results will apply, or can be adapted to other linear recurrences, such as the $(+,\times)$ linear systems of classical control theory.
According to the preliminary remarks above, an application of the results of this section will therefore be an algorithm to compute the makespan of the greedy execution of some URE. Note that this algorithm itself will not be greedy.
We shall first introduce some concepts and notations. We will then present the optimization problem which arises in the parallel computation of $(\max,+)$ URE, and apply the preceding results to solve it. Finally, we shall mention some particularities of stochastic versions of $(\max,+)$ systems.

2.7.1 Introduction

From now on, we therefore restrict our attention to "Linear Max-Plus Recurrences" (MPR), which are URE of the form:

$$X_i(n) = \max_{(j,\gamma)\in\mathcal{D}_i} (X_j(n-\gamma)+\sigma_{i,j,\gamma}),\ \sigma_{i,j,\gamma}\in\mathbb{R}^+ . \tag{2.8}$$

The assumption that $\sigma_{i,j,\gamma}$ is nonnegative is not necessary but natural because of the physical interpretation we gave above.
Let us introduce some definitions and notation.

Definition 2.7.1. *The* (max,+) semi-ring $\mathbb{R}_{\max}$ *is the set* $\mathbb{R}\cup\{-\infty\}$*, equipped with* max*, written additively (*i.e. $a\oplus b = \max(a,b)$*) and the usual sum, written multiplicatively (*i.e. $a\otimes b = a+b$*). The zero element is noted $\varepsilon=-\infty$, and the unit element is noted $e=0$.*

For matrices of appropriate sizes, we define $(A\oplus B)_{ij} = A_{ij}\oplus B_{ij} = \max(A_{ij},B_{ij})$, $(A\otimes B)_{ij} = \bigoplus_k A_{ik}\otimes B_{kj} = \max_k(A_{ik}+B_{kj})$, and for a scalar a, $(a\otimes A)_{ij} = a\otimes A_{ij} = a + A_{ij}$. When no confusion is possible, we abbreviate $A\otimes B$ to AB.

We can rewrite Equation (2.8) with the previously defined notations. Let $X(n)$ be the column vectors of coordinates $X_i(n)$ and let $A(\gamma)$ be the matrix with coordinates $A(\gamma)_{ij} = \sigma_{i,j,\gamma}$ if $(j,\gamma)\in\mathcal{D}_i$ and $A(\gamma)_{ij}=\varepsilon$ otherwise. We have

$$X(n) = A(0)\otimes X(n)\oplus A(1)\otimes X(n-1)\oplus\cdots\oplus A(\Gamma)\otimes X(n-\Gamma) , \tag{2.9}$$

where Γ is the maximum of the delays appearing in the sets $\mathcal{D}_i$.
This algebraic formulation enables some simple transformations. Let us define

$$A(0)^\star = \bigoplus_{n=0}^{\infty} A(0)^n = \bigoplus_{n=0}^{k} A(0)^n ,$$

where k is the size of matrix $A(0)$ and where $A(0)^0 = I$ is the identity matrix defined by $I_{ii} = e$ and $I_{ij} = \varepsilon, i \neq j$. It is easy to prove that $A(0)^\star$ is the formal inverse of $I - A(0)$, i.e. $A(0)^\star(I - A(0)) = (I - A(0))A(0)^\star = I$. Hence, Equation (2.9) can be transformed into:

$$X(n) = A(0)^\star A(1)X(n-1) \oplus \cdots \oplus A(0)^\star A(\Gamma)X(n-\Gamma) \ . \tag{2.10}$$

Equation (2.10) is nicer, because it involves only strictly positive delays, and is therefore obviously constructive.

Dependence graph The dependence graph has a general form as presented in §2.2.1. Its only characteristic is that the functions on the nodes are "max" applied to all entries.

Reduced graph A formalism naturally associated with (max, +) recurrences it that of *Petri nets*. The graphical formalism of Petri nets is close but different from the one we used for reduced graphs.
Petri nets consist of *transitions*, usually interpreted as service centers (processing units, etc.), and *places* containing *tokens*, usually interpreted as entities (programs, customers...) receiving services from transitions. Places are connected to transitions and transitions to places with directed arcs. It is therefore natural to speak of "input places", "output places" and so on. The passage from the reduced graph associated with a MPR to the corresponding Petri net consists in replacing nodes with transitions, and arcs with delay δ with a place containing δ tokens, connected to the corresponding transitions. Values of the delay therefore correspond to positions of tokens, called *markings*.
Figure 2.17 shows such a transformation.

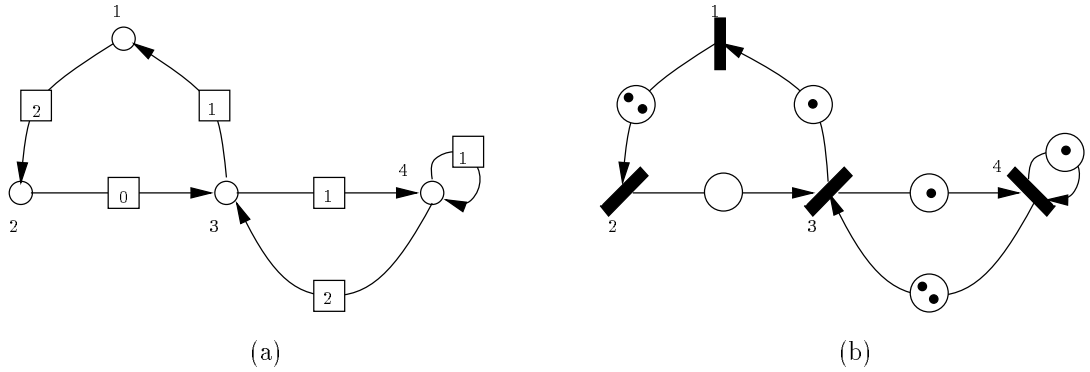


Figure 2.17: Transformation of a reduced graph (a) into a Petri net (b).

The Petri nets corresponding to reduced graphs have the particular property that places have exactly one input transition and one output transition. This property defines the class of *Event Graphs* (EG).
Petri nets are dynamical systems, in which tokens may move, according to the following rule. Transition may *fire*, thus removing one token from every input places and creating one in every output place. A fundamental remark is that firings in a Petri net are equivalent to retimings in

reduced graphs (see §2.5.1). Indeed, the equations describing the transformations of the marking after a firing are precisely of the form $\gamma' = \gamma + r(i) - r(j)$.
It is important to note that the usual convention for Petri nets is that a firing may occur only when at least one token is present in every input place of the transition. This requirement is dropped here. We therefore allow *negative markings*, which correspond to the negative delays of §2.5.
In *timed* event graphs, durations are associated with firings. Linear $(\max, +)$ systems of the form (2.8) describe the evolution of the associated event graphs in the following way. The variable $X_i(n)$ represents the instant at which the n-th firing of transition i starts, given that transitions start firing *as soon as possible*, that is, as soon as all tokens necessary are present in the input places *and* available, *i.e.* not involved in another firing. This interpretation holds under the assumption that the system is recycled, because the fact that tokens are involved in at most one firing implies that tokens go through transitions in a *first-in-first-out* (FIFO) order. Therefore, there are no overtaking of tokens, and the n-th firing of transition i requires the $(n - \gamma_{j,i})$-th token produced by transition j.

2.7.2 MPR of order 1

A standard step in the analysis of linear systems is the transformation of recurrences of order Γ such as (2.10) into an "equivalent" system of order 1. For general URE, this operation consists in introducing new variables $X_{k+1}, \ldots, X_\ell$ and new functions $G_i, i \in \{1, \ldots, \ell\}$ such that

$$X_i(n) = G_i(X_j(n - \gamma)), 1 \leqslant i \leqslant \ell,\ (j, \gamma) \in \mathcal{D}_i,\ n \in \mathbb{N}\ , \quad \gamma \leqslant 1\ . \tag{2.11}$$

This is usually done by setting $X_{(i-1)\times\Gamma+\gamma}(n) = X_i(n - \gamma)$, for $1 \leqslant i \leqslant k$ and $1 \leqslant \gamma \leqslant \Gamma$. The new number of variables is therefore $\ell = k\Gamma$.
In the case of $(\max, +)$ linear systems, the equivalent system of order 1 is characterized by a matrix of size $\ell \times \ell$. The recurrence becomes

$$X(n+1) \ = \ A(n) \otimes X(n)\ . \tag{2.12}$$

In some practical applications, it may be desirable to reduce this size as much as possible. An instance of such applications is described in the following section.

2.7.3 Parallel simulation of time varying MPR

For the purposes of this section, we informally introduce a generalization of the URE model (2.2), in which the functions F_i may additionally depend on n. The particular example we have in mind is that of MPR of the form (2.8) in which the numbers $\sigma_{i,j,\gamma}$ are allowed to depend on n. In the analysis of discrete event systems, these sequences are commonly assumed to be random. Consider therefore an URE defined with a sequence of (possibly random) functions :

$$\{(F_1^n, \ldots, F_k^n),\ n \in \mathbb{N}\}\ .$$

We consider the associated dependence graph $\mathcal{D}$ and the pebble game under the set of rules $\mathcal{M}_3$. Executions under this rule are regular, and therefore characterized by a finite set $\mathcal{A}(0) \subset \{1, \ldots, k\} \times \mathbb{Z}^-$ providing the position of the pebbles at step 0. Let $X(n), n \in \mathbb{N}$, be the vector

whose coordinates are $X_i(p+n)$ for $(i,p) \in \mathcal{A}(0)$. By definition of rule $\mathcal{M}_3$, see §2.3.2, we have that there exists a sequence of functions $\{\phi^n, n \in \mathbb{N}\}$ such that:

$$X(n+1) = \phi^n(X(n)) = \phi^n \circ \phi^{n-1} \circ \cdots \circ \phi^0(X(0)) \,. \tag{2.13}$$

The *simulation* of the system consists in computing the value of all $X(n)$. A possible algorithm for doing this with a parallel computer uses the so-called parallel prefix principle. Using the fact that the composition of functions is associative, it is possible to divide the computation of $\phi^n \circ \cdots \circ \phi^0$ in smaller products $\phi^p \circ \cdots \circ \phi^q$ which may be computed by different processors. Note that for this, it is necessary that the operators ϕ possess a numerical representation on which composition may be performed. This is typically the case for linear operators, which are represented by matrices, for which composition is equivalent to the common product. The parallel prefix algorithm therefore directly applies to the parallel simulation of MPR (2.12).
A way to quantify the efficiency of the parallel algorithm is to evaluate its PRAM complexity. It can be shown that the number of operations required to simulate the linear system up to time N with P processors is $\mathcal{O}(\ell^3(N/P + \log(P)))$, where ℓ is, as above, the size of the matrix characterizing the linear system.
In order to minimize the complexity of this algorithm, it is therefore necessary to find a representation of the MPR of minimal size.

2.7.4 Optimization results

We shall show in this section that finding the minimum possible size for an order 1 representation of a MPR can be done in polynomial time with respect to k and Γ. This problem appears to be new in the context of event graphs. Some preliminary results, which correspond to our mode A for counting memories may be found in [28].
The basic idea is that, given a marking of the event graph, it is possible to transform this graph by adding new transitions and places, in such a way that the resulting event graph, restricted to the original transitions, has the same dynamic behavior as the original one, and moreover, the marking of the places is less than one (see [73] for further discussions).
To see this, recall the discussion of §2.5.2 on counting the delays, and in particular Figure 2.13 (B). Interpreted in terms of Petri nets, this construction amounts to "factor out" tokens introducing dummy transitions in a tree-like fashion, as in Figure 2.18. The dummy transitions are assumed to have a firing time of 0, and are recycled (this is not shown on the figure).
The number of transitions in the resulting event graph is Γ_B. The MPR associated with the new graph has a maximum delay $\Gamma = 1$ and by (2.10), it has the desired order 1 form with $\ell = \Gamma_B$
This is already an improvement on the standard representation, but the results of §2.5 allow to improve this, by finding first an optimal marking of the net, that is, a marking such that the above transformation provides an event graph with $\ell = \min_{\mathcal{RE}} \Gamma_B$ or $\ell = \min_{\mathcal{NRE}} \Gamma_B$ transitions. It is indeed necessary to distinguish the two cases, according to whether negative markings are desirable or not. This can be understood as follows.
Assume that the marking corresponding to $\ell = \min_{\mathcal{RE}} \Gamma_B$ is negative. The event graph can be transformed into another equivalent one in the same way as for circuits in 2.6.3. The newly created transitions will then have firing times $\sigma_i(n)$ *equal* to some of the $\sigma_j(n+\gamma), \gamma > 0$ of the original transitions. Therefore, the construction of matrices $A(n)$ and $A(n+\gamma)$ in recurrence

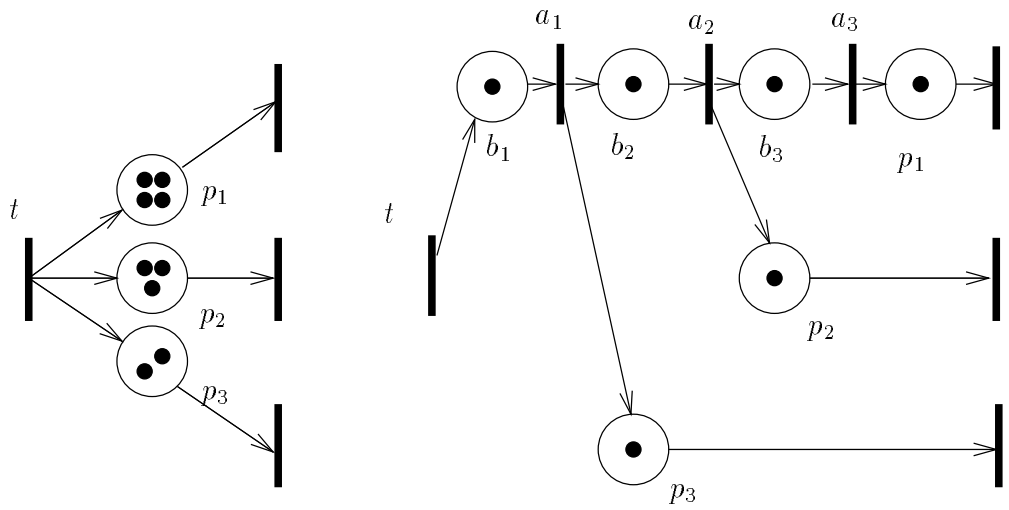


Figure 2.18: Forward transformation of event graphs.

(2.12) will use the same numbers σ. When two matrices do not use the same numbers, they are called *disjoint* matrices

In the parallel computing context, where the computations using $A(n)$ and $A(n+\gamma)$ are (possibly) done by different processors, it may be acceptable to use matrices which are not disjoint. For instance if $\sigma_i(n)$ does not depend on n, or if it can be computed in a deterministic way by the different processors.

However, it is not acceptable if the variables $\sigma_i(n)$ are *independent and identically distributed* (i.i.d.). In this case, we want the random matrices $A(n)$ and $A(n+\gamma)$ to be independently generated by their respective processor. This situation is the most common one in the context of discrete event system modeling, which we have already mentioned. It requires that all the matrices are disjoint.

In both cases, the optimal marking is found in polynomial time:

- If negative markings are acceptable, use the results of §2.4.3 to find a minimal consecutive cut.
- If not, use the algorithm of §2.6.4 to find a minimum compatible cut.

Remark 2.7.2. The optimality of the size of representation should be understood as the best possible that can be obtained without making assumptions on the value of the numbers $\sigma_{i,j,\gamma}(n)$. When these numbers are constant and known, this knowledge may be exploited to obtain a minimal representation in the sense of linear system theory [67], which is normally better than ours. A deeper investigation of the relations between the two approaches is an interesting direction for further research.

2.7.5 Backward transformation

To conclude the section, we make the remark that there are actually two ways to perform graph transformations : the *forward* transformation on downstream places as above (Figure 2.18), or the *backward* transformation on upstream places (Figure 2.19).

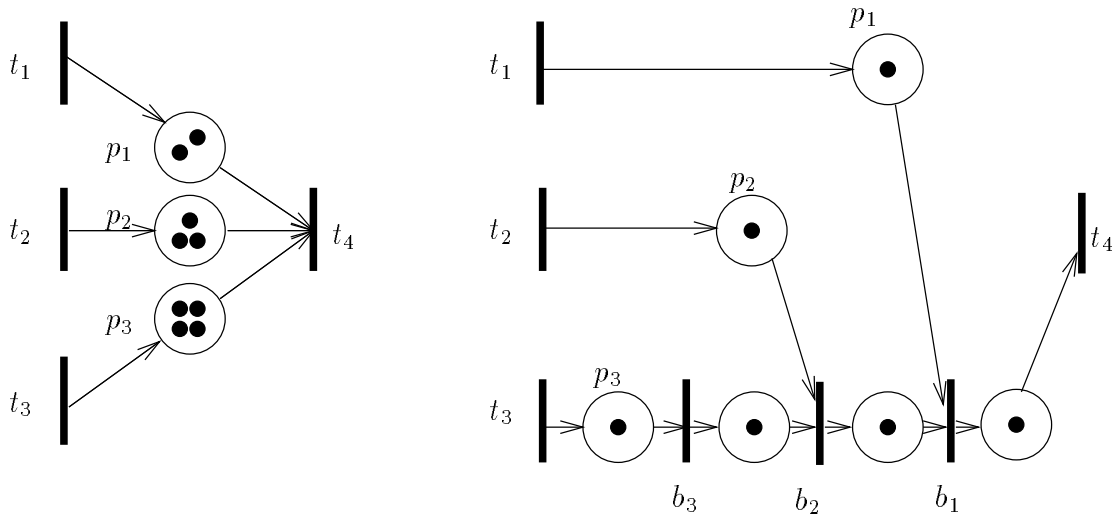


Figure 2.19: Backward transformation of event graphs.

Remark 2.7.3. It is important to note that the backward transformation is not possible for a general system of URE. A system of URE associated with the example of Figure 2.19 (before transformation) is of the form :

$$X_4(n) = F_4(X_1(n-2), X_2(n-3), X_3(n-4)) .$$

If we perform the backward transformation, the form of an URE associated with the new graph has to be :

$$X_4(n) = F(\ G(\ X_1(n-2), H(\ X_2(n-3), I(\ X_3(n-3)\ )\ )\ )\ ) ,$$

for some functions F, G, H and I associated with nodes t_4, b_1, b_2 and b_3 respectively. In general it is not possible to perform such a factorization of the original F_4 function. It becomes possible in MPR because of the special form of the functions F_i which are involved.

In the context of event graphs, the backward transformation is interesting since it results in a graph which might be smaller than the one obtained with the forward transformation.
The example displayed in Figure 2.20 has the following property. If we apply a forward transformation (after optimal legal retiming) the number of transitions is 7. However, a backward transformation yields a graph with only 6 transitions. It is also interesting to note that in this example, the size of the minimal (non compatible) cut in the dependence graph is 5. Therefore, this is an example where one can find an intermediate system of size 6, (strictly between the size of the smallest compatible cut and the size of the smallest cut) which allows a representation of

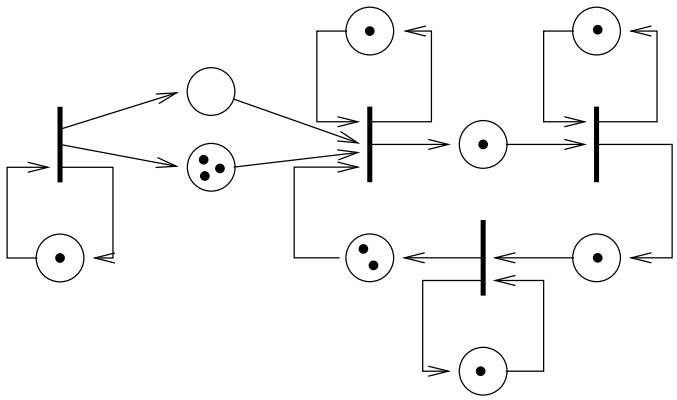


Figure 2.20: Optimal forward and backward transformations of this event graph yield to systems of different size.

the original system with disjoint matrices of size 6.

Determining an optimal backward transformation can be done using the previous results. In fact, it corresponds to the optimal forward transformation on the *reversed* event graph obtained by reversing the direction of all the arcs.

More generally, it is easy to come up with examples where the optimal transformation of an event graph involves both forward and backward transformations. Finding an algorithm to compute such an optimal mixed transformation is an interesting open problem.

2.7.6 Stochastic issues

The retimings used above for deriving optimal representations necessitate changes in the initial condition and shifts in the indices of the sequences $\{\sigma_{i,j,\gamma}(n)\}$.
When these sequences are random, these changes may be unnecessary, depending on the performances that are measured on the system.
Indeed, it is proved in Chapter 8 that *stationary* statistics of the MPR (such as asymptotic growth rate, and limit distributions for finite differences) are insensitive to the initial conditions under minimal stochastic assumptions on the sequences.

Chapitre 3

Graphical Approach of the Spectral Theory in the (max,+) Algebra

Approche Graphique de la Théorie Spectrale dans l'Algèbre (max,+)

Dans ce chapitre, on étudie plus particulièrement les systèmes (max,+) linéaires déterministes de dimension 3. On propose un nouvel outil de description du comportement spectral. Il s'agit de la représentation graphique des vecteurs propres et des domaines d'attraction dans un "espace projectif additif".

Ce chapitre est une adaptation de l'article [104]. Une version courte du même article paraîtra en octobre 95 dans *IEEE Transactions on Automatic Control* [106]. Cet article a grandement profité de nombreuses suggestions de François Baccelli et de Stéphane Gaubert. Sa présentation a également été améliorée grâce aux remarques de Damien Artiges et d'un rapporteur anonyme.

In this paper, we study matrices in the (max,+) algebra. We introduce a new tool for describing the deterministic spectral behaviour of matrices of size 3×3. It consists of a graphical representation of eigenvectors and domains of attraction in the projective space.

3.1 Introduction

Discrete Events Dynamic Systems (DEDS's) are a common framework to represent communication or manufacturing networks. Petri Nets, and more precisely Event Graphs, are an example of a formalism to study DEDS's. Event Graphs model phenomena such as synchronization or blocking. They have a simple interpretation in a nonconventional algebra, the (max,+) algebra. The spectral theory of matrices in the (max,+) algebra is now well known. It can be tracked back to Cuninghame-Green [49], Gondran and Minoux [77] or, for the Russian school, to Romanovskiĭ [123]. One of the main differences with the classical spectral theory is that there is a unique eigenvalue for irreducible matrices. As a consequence, the main interest and difficulty is to study eigenvectors associated with the unique eigenvalue. For a timed Event Graph, the eigenvalue is exactly the mean cycle time (inverse of the throughput rate). On the other hand, eigenvectors are associated with quantities such as : number of tokens in a place, waiting times or idle times. Multiple eigenvectors mean multiple possible regimes for these quantities.
In this paper, we present the classical spectral results under a new light. We develop a tool for describing the spectral behaviour of matrices of size 3×3. It consists of a graphical representation of asymptotic regimes in a projective space.

The paper is organized as follows. In Section **3.2**, we define the (max,+) algebra. In Section **3.3**, we propose, as an illustration, an example of a manufacturing model. In Section **3.4**, we review some results on the spectral theory in the (max,+) algebra. In Section **3.5**, we present also a complete spectral analysis of matrices of size 3 with the help of the graphical representation mentioned before. Sections **3.6** and **3.7** are devoted to applications of the graphical representation.

3.2 The (max,+) Algebra

We consider systems whose dynamic behaviour is driven by a recursive equation of the form :

$$x_i(n+1) \quad = \quad \max_{1\leqslant j\leqslant k}(A_{ij} + x_j(n)),\ i = 1, \ldots, k\,. \tag{3.1}$$

We allow A_{ij} to be equal to $-\infty$. Let us introduce some new notations.

Definition 3.2.1 ((max,+) algebra). *We consider the semiring $(\mathbb{R}\cup\{-\infty\}, \oplus, \otimes)$. The law $\oplus$ is "max" and $\otimes$ is the usual addition. We set $\varepsilon = -\infty$ and $e = 0$. The element ε is neutral for the operation $\oplus$ and absorbing for $\otimes$. The element e is neutral for $\otimes$. The law $\oplus$ is idempotent, i.e. $a \oplus a = a$. $(\mathbb{R}\cup\{\varepsilon\}, \oplus, \otimes)$ is an idempotent semiring or dioid. It is usually referred to as the (max,+) algebra (although it is not an algebra !, see [8], p.214). We shall denote it by $\mathbb{R}_{max}$.*

In the rest of the paper, the notations "+,×" will stand for the usual addition and multiplication. We will write ab for $a \otimes b$, however, whenever there is no possible confusion. For example, for $a \in \mathbb{R}$, $a^d = a^{\otimes d} = d \times a$.

We define the product spaces $\mathbb{R}_{max}^k$, $\mathbb{R}_{max}^{k\times k}$. We define the product of a vector by a scalar : $a \in \mathbb{R}_{max}, u \in \mathbb{R}_{max}^k$, $(a \otimes u)_i = a \otimes u_i = a + u_i$.
The matrix product is defined in a natural way, replacing $+$ and $\times$ by $\oplus$ and $\otimes$ respectively. Let $A, B \in \mathbb{R}_{max}^{k\times k}$,

$$(A \otimes B)_{ij} = \max_l (A_{il} + B_{lj}) = \bigoplus_l A_{il} \otimes B_{lj} \quad .$$

The matrix-vector product is defined in a similar way.

With these notations, the basic evolution Equation (3.1) takes a very simple and convenient form. It can be rewritten as :

$$x(n+1) \quad = \quad A \otimes x(n) \; . \tag{3.2}$$

Here $x(n) = (x_1(n), x_2(n), \ldots, x_k(n))'$ and A is a $k \times k$ matrix.

We consider the following "eigenvalue problem". We want to find nontrivial solutions of the equation :

$$A \otimes u = \lambda \otimes u \; , \tag{3.3}$$

where $A \in \mathbb{R}^{k\times k}$ is an irreducible (see Def. 3.4.2) matrix, u is a column vector (the "eigenvector") and λ is a real constant (the "eigenvalue"). We also define periodic solutions of the eigenvalue problem.

Definition 3.2.2. *A periodic solution (or regime) of period d is a set of vectors $\{u^1, \ldots, u^d\}$ of $\mathbb{R}^k$ verifying $Au^i = \lambda u^{i+1}$, $i = 1, \ldots, d-1$ and $Au^d = \lambda u^1$. It implies that the vectors $u^1, \ldots, u^d$ are eigenvectors of A^d.*

The eigenvalue of a matrix A gives the asymptotic growth rate of A^n/n (see Theorem 3.4.7 for a more precise statement). On the other hand, eigenvectors and periodic regimes are related with the problem of computing differences such as $A^{n+1}u - A^n u$ or $(A^n u)_i - (A^n u)_j$. These differences are related to many quantities of interest, see Sections 3.3 and 3.6. In this paper, we focus essentially on eigenvectors and periodic regimes.

3.3 A Simple Manufacturing Model

There are two types of items which have to be assembled together to form a part. There is a stock for each kind of item. We suppose that these stocks are infinite. Each time a part is completed at the assembly line, a new request is sent to the storage facilities. New items are then sent to the assembly line. We denote :

- α : operating time at the assembly line.
- $\beta_i, i = 1, 2$: communication time between the assembly line and stock i.
- $\gamma_i, i = 1, 2$: transportation time between the stocks and the assembly line.

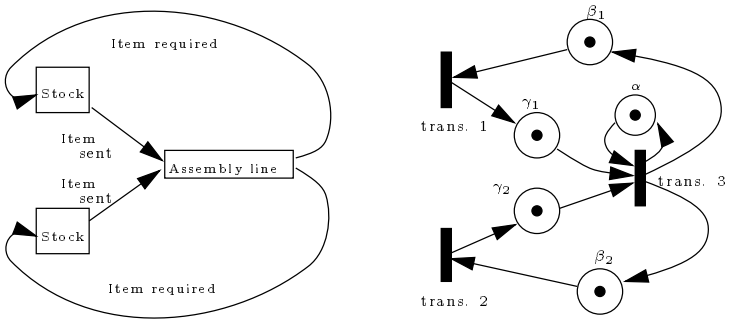


Figure 3.1: A manufacturing model and its Petri net representation.

We consider three daters $(x_i(n), i = 1, 2, 3)$ associated with this system. The first two correspond to the instants at which an item is sent from the stocks. The third one corresponds to the instants at which a part is completed at the assembly line. For example, $x_3(n)$ is the n-th instant of completion of a part at the assembly line. Then the (max,+) linear system corresponding to this system is the following one :

$$x(n+1) = M \otimes x(n),\ M = \begin{pmatrix} \varepsilon & \varepsilon & \beta_1 \\ \varepsilon & \varepsilon & \beta_2 \\ \gamma_1 + \alpha & \gamma_2 + \alpha & \alpha \end{pmatrix} . \tag{3.4}$$

The eigenvalue of M is $\lambda = \max(\alpha, (\beta_1 + \gamma_1 + \alpha)/2, (\beta_2 + \gamma_2 + \alpha)/2)$ (see Theorem 3.4.3). It corresponds to the mean cycle time, i.e. the inverse of the throughput of the system. One can also compute, for example, the idle time of the assembly line between the completion of a task and the beginning of the next one. Let us denote it by δ. We have :

$$\delta(n) = x_3(n) - x_3(n-1) - \alpha . \tag{3.5}$$

This example will be continued in Section 3.6.

This manufacturing system can be modeled using an Event Graph representation as shown in Figure 3.1. Event Graphs can efficiently model systems with synchronization, fork-join properties and/or blocking. It has been proved in Baccelli [4] that all Event Graphs can be described by an evolution equation of the form of Equation (3.1). For more insights on all modeling aspects, the reader is referred to Baccelli, Cohen, Olsder and Quadrat [8].

3.4 Spectral Theory in $\mathbb{R}_{max}$

The spectral theory of irreducible matrices in $\mathbb{R}_{max}$ is now classical. Most of the results have been proved by several authors independently. It makes it quite difficult to determine precise attributions. It seems that Theorem 3.4.3 is due to Cuninghame-Green [49]. It was also proved by Reiter [121] and Romanovskiĭ [123]. Versions of Theorem 3.4.6 were proved in [123] [50] and Gondran and Minoux [77]. Under the form proposed here, the result is from [44]. Theorem 3.4.7 is due to Cohen, Dubois, Quadrat and Viot [43] and [44]. For event graphs, a similar result was

proved by Chretienne [42]. A complete treatment of the spectral theory can be found in [8]. For the spectral theory of reducible matrices, the reader is referred to Gaubert [67] and Wende and al [139]. A spectral theory for non finite dimensions is proposed in Dudnikov [58]. However, the idea of illustrating the spectral behaviours by graphical representations in a projective space, as in Section 3.5, is new.

3.4.1 General presentation

We recall that we want to find non trivial solutions of the equation $Au = \lambda u$, where $A \in \mathbb{R}_{max}^{k\times k}$.

Definition 3.4.1. *The graph associated with a matrix A is a directed graph having a number of nodes equal to the size of A. It contains an arc from i to j iff $A_{ji} \neq \varepsilon$. The valuation of this arc is A_{ji}.*

Definition 3.4.2. *A matrix A is irreducible if : $\forall i,j \ \exists n \geqslant 0 \mid (A^n)_{ij} > \varepsilon$ (or equivalently if its graph is strongly connected). A matrix A is aperiodic if : $\exists n \geqslant 0, \ \forall i,j \mid (A^n)_{ij} > \varepsilon$.*

From now on, we consider only irreducible matrices in $\mathbb{R}_{max}^{k\times k}$.

Theorem 3.4.3. *For each circuit of the graph of A, $\zeta = \{t_1, t_2, \cdots, t_j, t_{j+1} = t_1\}$, we define its average weight by :*

$$p(\zeta) = (A_{t_1 t_j} \otimes \cdots \otimes A_{t_3 t_2} \otimes A_{t_2 t_1})/j \ ,$$

(here the division is the conventional one). Matrix A has a unique (non ε) eigenvalue, λ. It satisfies the relation $\lambda = \max_\zeta \ p(\zeta)$, where the maximum is taken over all the circuits of (the graph of) A.

There might be several eigenvectors. An eigenvector has all its coordinates different from ε (due to the irreducibility assumption).

Definition 3.4.4. *We normalize a matrix by dividing (in $\mathbb{R}_{max}$, i.e. by subtracting in the conventional algebra) all its entries by its eigenvalue.*

A normalized matrix has e as eigenvalue. Eigenvectors and periodic regimes are invariant by a translation of all the entries of a matrix by the same real constant. In the rest of the paper, we will write the matrix we want to study in a positive form (i.e. with all terms $> e$) or in a normalized form depending on which one seems more convenient.

Definition 3.4.5. *For a matrix A, with eigenvalue λ, we define :*

Critical circuit *A circuit ζ of A is said to be critical if its average weight is maximal, i.e. if $p(\zeta) = \lambda$.*

Critical graph *It consists of the nodes and arcs of A belonging to the critical circuit(s). A critical column (resp. line) of A is a vector $A_{.i}$ (resp. $A_{i.}$) where i belongs to the critical graph. A critical term of A is a term A_{ij} where i and j belong to the critical graph.*

For a general graph, we define :

Cyclicity *The cyclicity of a strongly connected graph is the greatest common divisor of the lengths of all the circuits. The cyclicity of a general graph is the least common multiple of the cyclicities of its maximal strongly connected subgraphs.*

To study the spectral behaviour of a matrix A, it is enough to know :

- The number of maximal strongly connected subgraphs (**s.c.s.**) of its critical graph.
- The cyclicity of its critical graph.

In the following, a matrix is said to be of type **SCS**j**-CYC**k if its critical graph has j s.c.s. and a cyclicity of k.

The two following theorems justify the previous assertion. For a normalized matrix A of size k, we define $A^+ = A \oplus A^2 \oplus \cdots \oplus A^k$. We check that $A^+ \oplus A^{k+1} = A^+$. We check also that A^+ has the same critical columns (resp. lines) as A.

Theorem 3.4.6. *Let A be a normalized matrix.*

a. Critical columns $A^+_{.i}$ are eigenvectors.

b. For i, j belonging to the critical graph, $\pi(A^+_{.i})$ and $\pi(A^+_{.j})$ are different iff i and j belong to two different s.c.s. of the critical graph.

c. Every eigenvector of A writes as a linear combination (in $\mathbb{R}_{max}$, see Section 3.4.2) of critical columns $A^+_{.i}$.

Because of c., the vectors $A^+_{.i}$, i belonging to the critical graph, are called the extremal eigenvectors.

Theorem 3.4.7. *For an irreducible matrix A of size k and whose eigenvalue is λ, there exist integers d and N such that :*

$$\forall n \geqslant N, \quad A^{n+d} = \lambda^{\otimes d} \otimes A^n \ .$$

Furthermore the smallest d verifying the property is equal to the cyclicity of the critical graph of A. From now on, we will call it the cyclicity of A. A cyclicity of d will provide periodic regimes of period d for the eigenvalue problem.

The good interpretation is that there exists an initial transient regime for the powers of a matrix A. After the transient regime, the sequence $\{A^n\}$ becomes periodic (more rigorously, it is the sequence $\{\pi(A^n)\}$ which becomes periodic, see Definition 3.4.9).
The term A^n_{ji} can be interpreted as the heaviest path of n steps starting from i and arriving at j in the graph of A. Theorems 3.4.3 and 3.4.7 state that the asymptotic growth rate of A^n is given by the circuits of A having the maximal average weight.

Definition 3.4.8. *Let A be an irreducible matrix and N be the smallest integer such that $\forall n \geqslant N$, $A^{n+d} = \lambda^{\otimes d} \otimes A^n$ (see Theorem 3.4.7). The matrices $\{A, \ldots, A^{N-1}\}$ constitute the so-called transient regime and the matrices $\{A^n, n \geqslant N\}$ the stationary regime. We say that a matrix $A^n, n \geqslant N$ is a stationary matrix associated with A, or the stationary version of A.*

3.4.2 The projective space

A "linear combination" (in $\mathbb{R}^k_{max}$) of eigenvectors is an eigenvector, i.e. if u_1 and u_2 are eigenvectors and $\alpha_1, \alpha_2 \in \mathbb{R}$, then $(\alpha_1 \otimes u_1) \oplus (\alpha_2 \otimes u_2)$ is also an eigenvector. In particular, if u is an eigenvector and $\alpha \in \mathbb{R}$, then $\alpha \otimes u$ is also one. This motivates the introduction of an "additive" projective space $\mathbb{PR}^k$.

Definition 3.4.9 ($\mathbb{PR}^k$). *The "additive" projective space $\mathbb{PR}^k$ is defined as the quotient of $\mathbb{R}^k$ by the parallelism relation :*

$$u, v \in \mathbb{R}^k \quad u \simeq v \Longleftrightarrow \exists a \in \mathbb{R} \text{ such that } u = a \otimes v \, .$$

Let π be the canonical projection of $\mathbb{R}^k$ into $\mathbb{PR}^k$.

The projection π can be interpreted geometrically. It is the orthogonal projection on the hyper-space orthogonal to the vector $\mathbb{1} = (1, \ldots, 1)'$. The projective space $\mathbb{PR}^k$ is isomorphic to $\mathbb{R}^{k-1}$. For irreducible matrices of size 2 or 3, we can represent, in $\mathbb{R} \simeq \mathbb{PR}^2$ and $\mathbb{R}^2 \simeq \mathbb{PR}^3$ respectively, eigenvectors and periodic regimes modulo the parallelism relation.

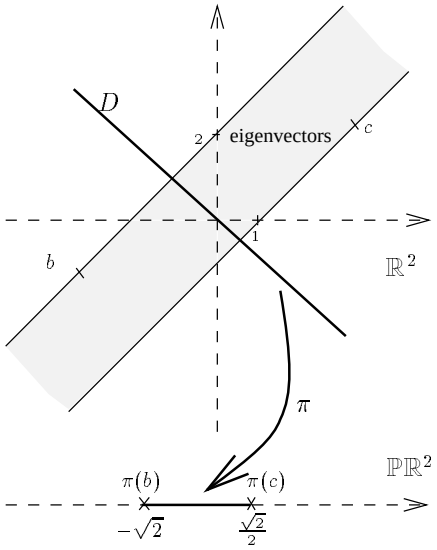


Figure 3.2: The projective space $\mathbb{PR}^2$.

We illustrate this on Figure 3.2 for the matrix :

$$R = \begin{pmatrix} e & -2 \\ -1 & e \end{pmatrix} .$$

This matrix verifies $R^+ = R$. By Theorem 3.4.6, the extremal eigenvectors are $\pi(e, -1)'$ and $\pi(-2, e)'$ and the set of eigenvectors is the set of linear combinations of these two points. It is the strip represented in Figure 3.2. The line D, in Figure 3.2, is the hyper-space orthogonal to the first bisecting line. To obtain the set of eigenvectors of R in $\mathbb{PR}^2$, we consider the orthogonal projection on D.

Let us introduce a distance $d(.,.)$ on $\mathbb{PR}^k$ which we are going to call the projective distance.

Definition 3.4.10 (projective distance). *We consider $x, y \in \mathbb{PR}^k$. Let $u, v \in \mathbb{R}^k$ be two representatives of x and y, i.e. $\pi(u) = x$ and $\pi(v) = y$.*

$$d(x,y) = d(u,v) = \bigoplus_i (u_i - v_i) \otimes \bigoplus_i (v_i - u_i) \ .$$

It is easy to check that $d(x,y)$ does not depend on the representatives u and v. It is also easy to check that it is a distance in $\mathbb{PR}^k$. It corresponds to the $\mathcal{L}_\infty$ distance on the projective space $\mathbb{PR}^k$, see Figure 3.3. We write either $d(x,y)$ or $d(u,v)$ with a little abuse of notation. We have the following property.

Proposition 3.4.11. *Let A be an irreducible matrix of size k. Let u, v be two vectors of $\mathbb{R}^k_{max}$. We have :*

$$d(Au, Av) \leqslant d(u,v) \ .$$

There is no simple criterion to get a strict inequality. For a proof of this result see Proposition 6.3.8. As an easy corollary of Theorem 3.4.7, we obtain that for an irreducible matrix A, there exists N such that $\forall u, v \in \mathbb{R}^k$, $\forall n \geqslant N$, $d(A^n u, A^n v) = d(A^N u, A^N v)$.

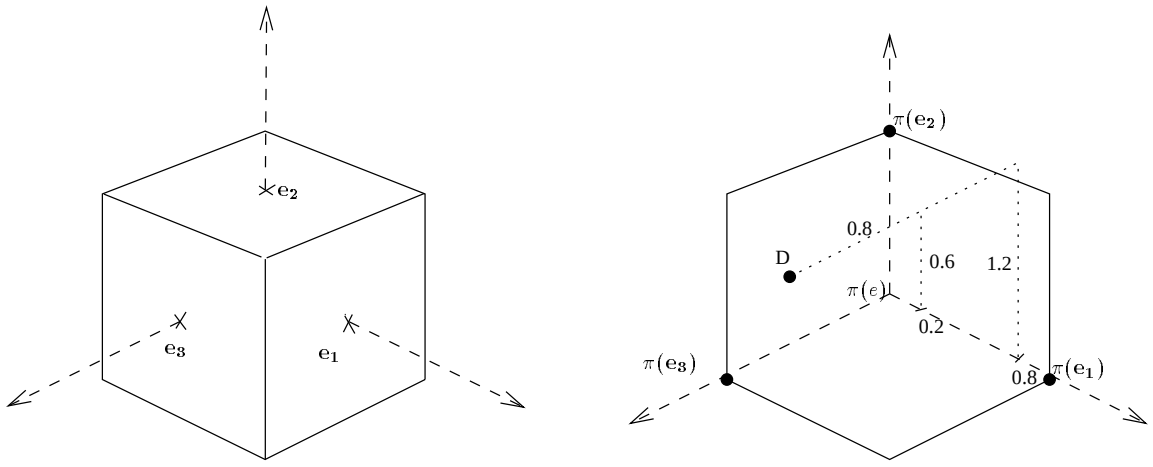


Figure 3.3: Unit ball of the projective distance in $\mathbb{PR}^3$.

Let us represent the unit ball of the projective distance in $\mathbb{PR}^3$. The regular hexagon in Figure 3.3 is the section of the unit square (i.e. the unit ball of $\mathbb{R}^3$ for the $\mathcal{L}_\infty$ norm) by the projection plane. The three represented axes are the orthogonal projection of the basis of $\mathbb{R}^3$. The represented points are $\pi(e) = \pi(e,e,e)'$, $\pi(e_1) = \pi(1,e,e)'$, $\pi(e_2) = \pi(e,1,e)'$, $\pi(e_3) = \pi(e,e,1)'$ and $D = \pi(0.2, 0.6, 0.8)'$.

The practical way of representing a point X of $\mathbb{PR}^3$ is to choose a vector ($\in \mathbb{R}^3$) in the parallelism class of X and to draw it in the three axes obtained by projection of the orthonormal basis of

$\mathbb{R}^3$ (it is easy to check that the point we obtain does not depend on the representative in the parallelism class). The point D of Figure 3.3 illustrates this. We have drawn two constructions : one corresponding to (0.2, 0.6, 0.8) and the other one to $(0.8,\ 1.2,\ 1.4) = 0.6 \otimes (0.2,\ 0.6,\ 0.8)$.

Let us illustrate what the "linear combination" of two vectors means in $\mathbb{PR}^k$. We consider examples of dimension 3. Let $u = (u_1, u_2, u_3)'$ and $v = (v_1, v_2, v_3)'$ be two vectors of $\mathbb{R}^3$. Let $\lambda, \mu \in \mathbb{R}$.

$$\pi(\lambda \otimes \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \oplus \mu \otimes \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}) = \pi(\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \oplus (\mu - \lambda) \otimes \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}) .$$

Let us assume for example that we have,

$$u_1 - v_1 \leqslant u_2 - v_2 \leqslant u_3 - v_3 .$$

Depending on the value of $\alpha = \mu - \lambda$, there are four possible cases.

1. If $\alpha \leqslant u_1 - v_1$, then $\pi(u \oplus \alpha v) = \pi(u)$.
2. If $u_1 - v_1 \leqslant \alpha \leqslant u_2 - v_2$, then $\pi(u \oplus \alpha v) = \pi(\alpha v_1, u_2, u_3)'$.
3. If $u_2 - v_2 \leqslant \alpha \leqslant u_3 - v_3$, then $\pi(u \oplus \alpha v) = \pi(\alpha v_1, \alpha v_2, u_3)'$.
4. If $u_3 - v_3 \leqslant \alpha$, then $\pi(u \oplus \alpha v) = \pi(v)$.

This particular example corresponds to the case of points $\pi(e_3)$ ($\pi(u)$) and $\pi(e_1)$ ($\pi(v)$) in Figure 3.3. The "line" between $\pi(e_3)$ and $\pi(e_1)$, in Figure 3.3, is the set of linear combinations of the two points. When two values are equal in $\{u_i - v_i,\ i = 1,2,3\}$, the picture is degenerate (see Figure 3.5, matrix D).
The regular hexagon of Figure 3.3 is the convex hull of the points $\pi(e_1), \pi(e_2)$ and $\pi(e_3)$, i.e. the set :

$$\{\pi(\alpha \otimes e_1 \oplus \beta \otimes e_2 \oplus \gamma \otimes e_3),\ \alpha, \beta, \gamma \in \mathbb{R}_{max}\}.$$

Let us denote by $\pi(f_1), \pi(f_2)$ and $\pi(f_3)$ the three other vertices of the regular hexagon. More precisely, $\pi(f_1) = \pi(e, 1, 1)'$, $\pi(f_2) = \pi(1, e, 1)'$ and $\pi(f_3) = \pi(1, 1, e)'$. One can check that the convex hull of these points is the union of the segments $[\pi(f_i), \pi(e)]$.

3.4.3 Change of basis

A matrix A of $\mathbb{R}_{max}^{k \times k}$ can be considered as a "linear" operator on $\mathbb{R}^k$. We want to have a formula of change of basis for the matrix associated with a given linear operator. We are only interested in permutation of the coordinates and translation of the origin.

Definition 3.4.12. *Let σ be a permutation of $\{1, \ldots, k\}$. The matrix of permutation associated with σ is P defined by :*

$$P_{\sigma(i),i} = e,\ P_{ji} = \varepsilon,\ \forall j \neq \sigma(i) .$$

Lemma 3.4.13. *Let A be a $k \times k$ matrix and let $\hat{A}$ be the matrix associated with the same operator in a new basis obtained from the original one by a permutation σ of the coordinates. Matrix P is the permutation matrix associated with σ and P^{-1} is the one associated with σ^{-1}. We have $\hat{A} = P^{-1} \otimes A \otimes P$.*

We consider a matrix A. We denote by $\tilde{A}$ the matrix associated with the same operator in a new basis obtained from the original one by a translation of the origin of the projective space.

Lemma 3.4.14. *Let A be a $k \times k$ matrix. Let $u \in \mathbb{R}^k$ be (a representative of) the new origin written in the old basis. In the new basis, we have $\tilde{A} = P^{-1} \otimes A \otimes P$, where $P_{ii} = u_i$, $P_{ij} = \varepsilon$, $\forall i \neq j$.*

Proof. Let $v = (v_1, \cdots, v_k)'$ be a vector written in the old basis and let $\tilde{v} = (\tilde{v}_1, \cdots, \tilde{v}_k)'$ be this same vector in the new basis. We have $\tilde{v}_i + u_i = v_i$. We set $Av = w$ and $w = (w_1, \cdots, w_k)'$ and $(\tilde{w}_1, \cdots, \tilde{w}_k)'$ in the new and the old basis respectively.

$$\begin{aligned}(\tilde{A}\tilde{v})_i = (P^{-1} \otimes A \otimes P\tilde{v})_i &= (P^{-1} \otimes Av)_i \\ &= (P^{-1}w)_i = \tilde{w}_i\end{aligned}$$

□

It might be interesting to get another intuition on what a change of origin means. We present now an interpretation suggested by the modeling of Stochastic Event Graphs. Let us consider the communication graph associated with a positive and irreducible matrix $A \in \mathbb{R}_{max}^{k \times k}$. We consider that there is a clock associated with each node of A. Let u be a vector of $\mathbb{R}^k$. We interpret u_i as a date of occurrence of a first event at node i. Then $(Au)_j$ is interpreted as the date of occurrence of the second event at node j. In this framework, a "change of origin" is just a change of the origin of time for some or all of the daters. It does not modify of course the evolution of the system.
The critical graph of a matrix is not modified by a change of basis. Lemma 3.4.15 shows that some bases have a particular interest.

Lemma 3.4.15. *We consider a matrix A, irreducible, of size k. Let λ be the eigenvalue and u an eigenvector of A. Let P be the matrix of change of the origin associated with u. Let $\tilde{A} = P^{-1} \otimes A \otimes P$. We have the following property $\forall i, j \in 1, \cdots, k$, $\tilde{A}_{ij} \leqslant \lambda$ and $\forall p, q$ such that (p, q) belongs to the critical graph (i.e. for all critical terms), we have $\tilde{A}_{qp} = \lambda$.*

Proof. We set $\mathbf{e} = (e, \ldots, e)'$.

$$\tilde{A}\mathbf{e} = \left(\bigoplus_j \tilde{A}_{1j}, \ldots, \bigoplus_j \tilde{A}_{kj} \right) .$$

But we also have that $\mathbf{e}$ is an eigenvector of $\tilde{A}$, $\tilde{A}\mathbf{e} = P^{-1}AP\mathbf{e} = P^{-1}Au = P^{-1}\lambda u = \lambda\mathbf{e}$. It implies that $\forall i$, $\bigoplus_j \tilde{A}_{ij} = \lambda$, which proves the first part of the lemma. Let us suppose there exist p, q such that (p, q) belongs to the critical graph and $\tilde{A}_{qp} < \lambda$. There is a critical circuit involving the arc (p, q). Using the first part of the lemma and $\tilde{A}_{qp} < \lambda$, we conclude that the mean weight of this critical circuit is strictly smaller than λ, which is a contradiction. □

3.4.4 Classification

The set of eigenelements of a matrix is the set of eigenvectors and periodic regimes. Each spectral behaviour corresponds to a specific form for the set of eigenelements. It is possible to prove that this set is compact iff the matrix is aperiodic (Def. 3.4.2). In the following, we restrict our attention to aperiodic matrices.

It is enough to study the sets of eigenelements for some *canonical* matrices of each spectral type scs$p$-cyc$l$. Let $A \in \mathbb{R}_{max}^{k\times k}$ be a given irreducible and aperiodic matrix. Its set of eigenelements can be obtained from the one of a canonical matrix by applying an homothety and a translation. Next algorithm propose a systematic way to determine the canonical matrix associated with A and to determine the homothety and the translation.

Algorithm

Step 1. *Find the eigenvalue and normalize A.*

Let λ be the eigenvalue of A. Normalize matrix A. For sake of simplicity, we keep the original notation, i.e. $A := A - \lambda$. The set of eigenelements is not changed by the normalization.

Step 2. *Find an eigenvector of A and write A in a new basis.*

Let u be an eigenvector of A. Consider a new basis obtained from the original one by a translation of u of the origin. For simplicity, we keep the notation A for the operator in the new basis. By Lemma 3.4.14, we have $A := P^{-1}AP$, where $P_{ii} = u_i$, $P_{ij} = \varepsilon, i \neq j$.
By Lemma 3.4.15, all critical terms of A are now equal to e and all non-critical terms are less than or equal to e.

Step 3. *Determine the spectral type of A.*

Compute the number of strongly connected subgraphs (p) and the cyclicity (l) of the critical graph. The spectral type of A is **scs$p$-cyc$l$**.

Step 4. *Compute the projective size of A.*

Compute the matrix A^l and the matrix $(A^l)^+$. Let $\mathcal{C}$ be the set of couples (i,j) such that $A_{.j}$ is a critical column (Definition 3.4.5). We set $\alpha = |\min_{(i,j)\in\mathcal{C}}(A^l)^+_{ij}|$. We call α the projective size of A. If $\alpha \neq 1$, we scale the matrix A. We set $A := A/\alpha$, each entry of A is divided (in the conventional algebra) by α.

Step 5. *Check non-critical terms of A. Final classification.*

Consider the terms $(A^l)^+_{ij}$ for the couples (i,j) such that $(i,j) \in \mathcal{C}$ and (i,j) is not a critical arc. If they are all equal to -1, matrix A is said to be regular. We say that the set of eigenelements of A has a *basic* form. If they are not all equal to -1, the set of eigenelements is modified and is said to be *non-basic*.

With this algorithm, we have associated with the original matrix a canonical matrix. Let S be the set of eigenelements of the canonical matrix. To obtain the set of eigenelements of the original matrix, it is necessary to apply to S :

1. An homothetic transformation of center $e = (e, \ldots, e)'$ and ratio α, where α is the projective size defined in Step 4.
2. A translation of u, where u is the eigenvector computed in Step 2.

Remark For some details on the practical implementation of this algorithm and for complexity results, see Chapter 5.

For matrices of dimension 3, a graphical representation of the sets of eigenelements of canonical matrices is proposed in §3.5.

3.5 Illustrated Spectral Theory in Dimension 3

We are ready to take a closer look at irreducible and aperiodic matrices of size 3. Using Theorems 3.4.6 and 3.4.7, it is easy to show that there are only six possible spectral behaviours, which can be sorted in four categories.

- scs1-cyc1
- scs2-cyc2
- scs3-cyc1 and scs1-cyc3.
- scs2-cyc1 and scs1-cyc2.

For each spectral type, we are going to draw the set of eigenelements, in $\mathbb{PR}^3$. We will also represent the domains of attraction of the different eigenelements. For a matrix A, we call domain of attraction of an eigenvector (resp. of a periodic regime) the set of initial conditions $\{x_0\}$ such that $\pi(A^n x_0)$ converges to that eigenvector (resp. periodic regime). By Theorem 3.4.7, this convergence occurs in finite time.

Regular sets of eigenelements have been represented for each spectral type, see Table 1. Examples of non-regular sets are also given for each spectral type, see Table 2.

Table **1.**

Spectral type	*Figure n°*
scs3-cyc1	3.4
scs1-cyc3	3.6
scs2-cyc1	3.8
scs1-cyc2	3.8,3.11
scs2-cyc2	3.12

, Table **2.**

Spectral type	*Figure n°*
scs3-cyc1	3.5
scs1-cyc3	3.7
scs2-cyc1	3.9
scs1-cyc2	3.9,3.10
scs2-cyc2	3.13

.

3.5.1 Scs1-Cyc1

Let A be a scs1-cyc1 matrix. We denote by v the unique eigenvector of A. By applying Theorem 3.4.7, we obtain :

$$\exists N, \forall n \geqslant N, \forall u_0 \in \mathbb{R}^k, \ \ \pi(A^n u_0) = \pi(v) \, .$$

The domain of attraction of $\pi(v)$ is $\mathbb{PR}^k$. This case is of special importance for stochastic models (see Chapter 6). There is no figure corresponding to this case as the spectral behaviour is trivial.

3.5.2 Scs3-Cyc1 and Scs1-Cyc3

If A is a scs1-cyc3 matrix, then A^3 is a scs3-cyc1 matrix (but the converse is false !). For example,

$$A = \begin{pmatrix} . & . & e \\ e & . & . \\ . & e & . \end{pmatrix}, \quad B = A^3 = \begin{pmatrix} e & . & . \\ . & e & . \\ . & . & e \end{pmatrix}, \tag{3.6}$$

where (.) stands for -1.

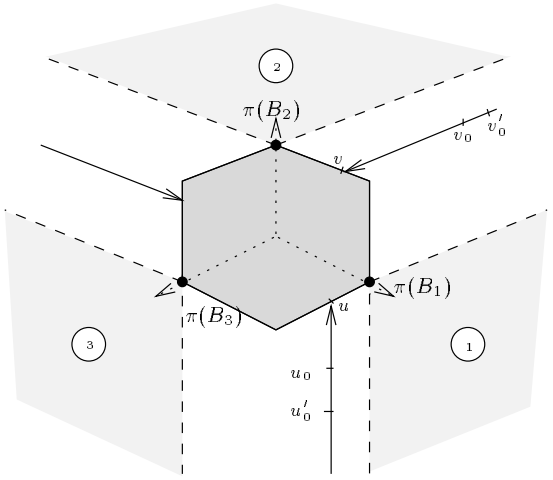


Figure 3.4: scs3-cyc1, set of eigenvectors of B.

We consider first the scs3-cyc1 case. There are three extremal eigenvectors and no periodic regime of period greater than one (Theorems 3.4.6 and 3.4.7). Let us consider more specifically the matrix B defined in (3.6). It is a normalized matrix and we check that $B^+ = B^2 = B$. By Theorem 3.4.6, the three columns B_1, B_2 and B_3 of B are the extremal eigenvectors. The set of eigenvectors is the $\mathbb{R}_{max}$ convex hull of these three eigenvectors. In Figure 3.4, it is represented in dark gray.
If the initial condition x_0 is in the light gray zone number i, then the limit value of $\pi(B^n x_0)$ is $\pi(B_i)$. If the initial condition is in one of the white strips, then the limit value is the nearest point for the projective distance (and this limit is attained in one step as $B^2 = B$). For example, for initial conditions u_0 or u_0' (resp. v_0 or v_0') the limit value is u (resp. v).

We will now consider what happens if we modify the non-critical terms of the matrix B. We consider three different examples to illustrate it.

$$C = \begin{pmatrix} e & . & . \\ -\frac{1}{2} & e & . \\ . & . & e \end{pmatrix}, \; D = \begin{pmatrix} e & -0.6 & . \\ . & e & 0.6 \\ . & . & e \end{pmatrix}, \; E = \begin{pmatrix} e & . & -\frac{1}{2} \\ -0.2 & e & . \\ -0.2 & -\frac{1}{2} & e \end{pmatrix},$$

where $(.) = -1$. We have represented the sets of eigenvectors of these matrices in Figure 3.5.

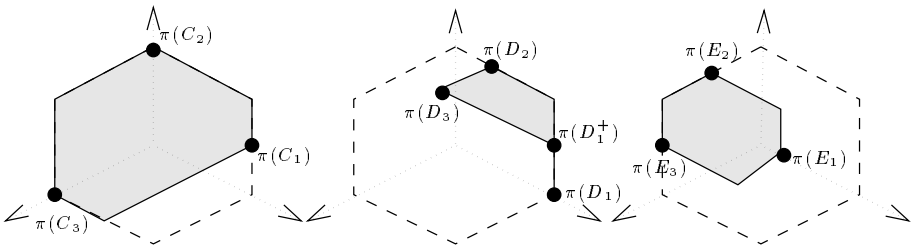


Figure 3.5: scs3-cyc1, sets of eigenvectors of C, D and E.

We can represent these sets using the procedure described in Section 3.4.2. Let us consider the matrix C for example. We represent the three columns of $C^+ = C$, $\pi(C_1)$, $\pi(C_2)$ and $\pi(C_3)$. The convex hull of these three points is the set of eigenvectors of C.

Now we consider the case of scs1-cyc3 matrices. There is only one eigenvector but there are periodic regimes of period 3. The set of periodic regimes of period 3 of a scs1-cyc3 matrix M is equal to the set of eigenvectors of M^3. Let us consider more specifically the matrix A defined in (3.6) above. It is easy to check that the unique eigenvector of A is $\mathbf{e} = (e, e, e)'$. We know the form of the set Π of periodic regimes of A (it is the set of eigenvectors of $A^3 = B$, see Figure 3.4). To go further, we want to characterize, given an initial condition $u \neq \mathbf{e}$ in the hexagon Π, the periodic regime associated with u. By Theorem 3.4.7, this periodic regime is $\{u, Au, A^2u\}$. We have $A^3u = u$ which implies $d(A^3u, \mathbf{e}) = d(u, \mathbf{e})$. By Proposition 3.4.11, we have that :

$$d(A^3u, \mathbf{e}) \leqslant d(A^2u, \mathbf{e}) \leqslant d(Au, \mathbf{e}) \leqslant d(u, \mathbf{e}) .$$

We conclude that :

$$d(A^2u, \mathbf{e}) = d(Au, \mathbf{e}) = d(u, \mathbf{e}) .$$

The points of a periodic regime are at a constant distance (for the projective distance) of the unique eigenvector $\mathbf{e}$. Furthermore, the three points $\{u, Au, A^2u\}$ must be invariant by a permutation of the three projective axes which characterize them completely. The direction of rotation depends on the critical circuit. It will be counter-clockwise if the critical circuit $(1, 2, 3)$ and clockwise if it is $(1, 3, 2)$. For example A and A^2 have the same set of periodic regimes but opposite directions of rotation.
We have represented in Figure 3.6, the set of eigenelements of matrix A. If the initial condition is in one of the gray zones, then the stationary periodic regime is $\{\pi(A_1), \pi(A_2), \pi(A_3)\}$. If the initial condition is in one of the white strips, the limit regime consists of three points on the boundary of the hexagon. For example for an initial condition u_0 or u_0', the limit regime is $\{u_1, u_2, u_3\}$. More precisely, we have :

$$\pi(Au_0) = \pi(u_1),\ \pi(A^2u_0) = \pi(u_2),\ \pi(A^3u_0) = \pi(u_3),\ \pi(A^4u_0) = \pi(u_1), \ldots ,$$

$$\pi(Au_0') = \pi(u_3),\ \pi(A^2u_0') = \pi(u_1), \ldots .$$

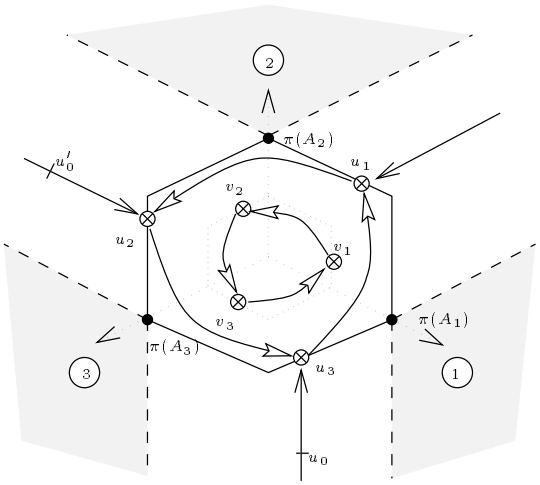


Figure 3.6: scs1-cyc3, set of periodic regimes of A.

If the initial condition u belongs to Π, the stationary periodic regime is $\{u, Au, A^2u\}$. We have also drawn an example of such a regime ($\{v_1, v_2, v_3\}$).

What happens if we perturb non-critical termsΓ To describe it, it will be useful to define the notion of sub-diagonals.

Definition 3.5.1. *Let M be a matrix of size k. We call i^{th} sub-diagonal of M the terms*

$$\{M_{i1}, M_{i+1,2}, \ldots, M_{i+k-i,1+k-i}, M_{1,2+k-i}, \ldots, M_{i-1,k}\} = \{M_{i-1+k,k\,[k]},\ \forall k\}\ .$$

For example, the first sub-diagonal is the diagonal of the matrix ! For the matrix A above, the critical sub-diagonal is the second one. If we increase a non-critical term (i.e. a term of the first or third sub-diagonal), after a transient regime, the whole sub-diagonal will be equal to this term. Let us consider an example.

$$A' = \begin{pmatrix} a & b_1 & e \\ e & . & . \\ b_2 & e & . \end{pmatrix} \longrightarrow\ (A')^4 = \begin{pmatrix} a & b & e \\ e & a & b \\ b & e & a \end{pmatrix},\ (A')^5 = \begin{pmatrix} b & e & a \\ a & b & e \\ e & a & b \end{pmatrix},\ \ldots,$$

with $(.) = -1$, $-1 < a, b_1, b_2 < e$, $b = b_1 \oplus b_2$[1] This provides us with specific pictures for the sets of periodic regimes. When we increase continuously a non-critical term, this set evolves in the same manner as the diaphragm of a camera. Let us illustrate it in Figure 3.7.

$$F = \begin{pmatrix} -0.8 & . & e \\ e & -0.8 & . \\ . & e & -0.8 \end{pmatrix},\ G = \begin{pmatrix} -0.5 & . & e \\ e & -0.5 & . \\ . & e & -0.5 \end{pmatrix},$$

[1]The projective size of matrix A' is here $\inf(a, b)$, see Stage 6 of the algorithm of Section 3.4.4.

$$H = \begin{pmatrix} -0.2 & . & e \\ e & -0.2 & . \\ . & e & -0.2 \end{pmatrix}, \ (.) = -1 \ .$$

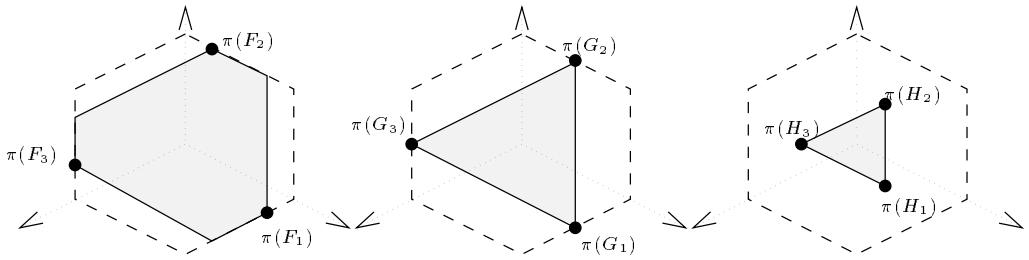


Figure 3.7: scs1-cyc3, sets of periodic regimes of F, G and H.

When the terms of the diagonal become equal to e, we obtain a scs1-cyc1 matrix with $\mathbf{e} = (e, \ldots, e)'$ as the unique eigenvector. When the terms of the diagonal become greater than e, then we get a scs3-cyc1 matrix for which the set of eigenvectors is similar to the one of Figure 3.4.

Remark In the cases we have been dealing with so far, domains of attraction had a very easy algebraic characterization. In fact for a matrix M and an initial condition u, the limit value of $\pi(M^n u)$ was the "nearest" (for the projective distance) eigenvector or periodic regime. This characterization will not be always true for the examples to come which makes the description of domains of attraction more delicate. For more insights, see Proposition 4.4.1.

3.5.3 Scs2-Cyc1 and Scs1-Cyc2

In the same way as previously, if M is a scs1-cyc2 matrix then M^2 is a scs2-cyc1 matrix, the converse being false. For example,

$$I = \begin{pmatrix} . & e & . \\ e & . & . \\ . & . & -2 \end{pmatrix}, \ J = I^2 = \begin{pmatrix} e & . & . \\ . & e & . \\ . & . & -2 \end{pmatrix}, \ (.) = -1 \ ,$$

Let us consider the scs2-cyc1 case and more precisely the matrix J. By Theorem (3.4.6), there are two extremal eigenvectors $\pi(J_1)$ and $\pi(J_2)$ (the first two columns of $J^+ = J^2 = J$) and no periodic regime of period greater than 1. In Figure 3.8, we have represented eigenvectors and domains of attraction for J.
There is a symmetry axis for the whole figure (corresponding to the fact that matrix J is unchanged by a permutation of the first two columns). The set of eigenvectors is given by the linear combinations of $\pi(J_1)$ and $\pi(J_2)$. As opposed to the scs3-cyc1 case, no eigenvector has a domain of attraction restricted to itself. If the initial condition x_0 is in the gray zones 1 or 2, the limit value of $\pi(J^k x_0)$ will be $\pi(J_1)$ or $\pi(J_2)$ respectively. If it is in zone 3, then the limit value will be $\pi(j) = \pi(e, e, -1)'$. When the initial condition is in one of the white strips, the limit value is given by the arrows. For example, if the initial condition is u_0, u_1 or u_2, the limit regime is u.

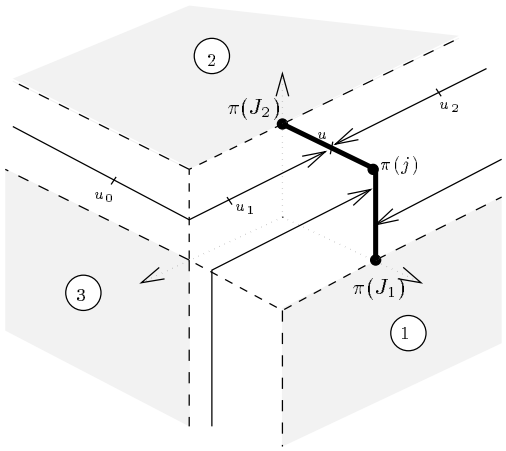


Figure 3.8: scs2-cyc1 (resp. scs1-cyc2), set of eigenelements of J (resp. I).

The picture remains the same for matrix I which is scs1-cyc2. There is only one eigenvector which is $\pi(j) = \pi(e, e, -1)'$. The "line" between $\pi(J_1)$ and $\pi(J_2)$ is the set of periodic regimes of period 2. Two points of this set belong to the same periodic regime if they are "symmetric" with respect to $\pi(j)$. For an initial condition in zone 3, the limit regime is the eigenvector $\pi(j)$. For an initial condition in zones 1 or 2, the limit regime is $\{\pi(J_1), \pi(J_2)\}$ and so on.

We want to analyze what happens if we modify non-critical terms. We have to distinguish between modifications of terms belonging to critical columns (columns 1 and 2 here) and of terms belonging to non-critical columns. If we modify a term belonging to a critical column, the set of eigenvectors (obtained as the convex hull of critical columns) will also be modified. On the other hand, it is possible that a modification of a term of the non-critical column does not affect the set of eigenvectors but only the domains of attraction. Let us illustrate this idea on Figure 3.9. It corresponds to the matrices :

$$K = \begin{pmatrix} e & . & -0.5 \\ . & e & . \\ . & . & -2 \end{pmatrix}, \; L = \begin{pmatrix} e & . & 0.5 \\ . & e & . \\ . & . & -2 \end{pmatrix}, \; (.) = -1 \, .$$

For matrix K, the set of eigenvectors is not modified, but the domains of attraction are. Figure 3.9 has to be interpreted in the same way as previously. The extremal eigenvectors are $\pi(K_1)$ and $\pi(K_2)$, the critical columns of K. The gray zones 1 and 2 are the domains of attraction of $\pi(K_1)$ and $\pi(K_2)$ respectively. If the initial condition u_0 is in zone 3, the limit value of $\pi(K^n u_0)$ will be $\pi(k) = \pi(1, 0.5, e)'$.

Matrix L is not stationary, see Definition 3.4.8. In fact, The stationary matrix associated with L is :
the domains of attraction and the set of eigenvectors are modified. In fact, The stationary matrix associated with L is :

$$L^2 = \begin{pmatrix} e & -0.5 & 0.5 \\ . & e & -0.5 \\ . & . & -0.5 \end{pmatrix}, \; (.) = -1 \, .$$

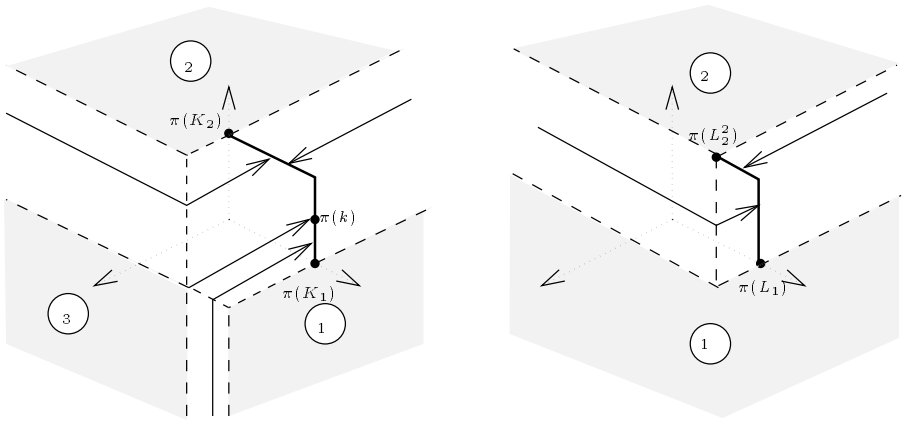


Figure 3.9: scs2-cyc1, sets of eigenvectors of K and L.

For matrix L^2, a term of a critical column has been modified. As a consequence, the domains of attraction **and** the set of eigenvectors are modified. The extremal eigenvectors $\pi(L_1^2)$ and $\pi(L_2^2)$ are the critical columns of matrix L^2. In this example, zones 1 and 3 have melted. They constitute the domain of attraction of $\pi(L_1^2)$.

Let us now consider what happens when we modify non-critical terms of a scs1-cyc2 matrix, M. The analysis made before remains valid. The set of periodic regimes of M is exactly the set of eigenvectors of the scs2-cyc1 matrix M^2. One interesting point to notice is that there might be no symmetry axis although all the stationary regimes are periodic of period 2. Figure 3.10 provides an example of this behaviour. It corresponds to the matrix :

$$O = \begin{pmatrix} . & e & . \\ e & . & . \\ -0.5 & . & -1.5 \end{pmatrix}, \ (.) = -1 \, .$$

Let $\pi(O_1^2)$ and $\pi(O_2^2)$ be the two extremal points of the set of periodic regimes (i.e. the critical columns of O^2, the stationary version of O). The unique eigenvector, $\pi(o) = \pi(e, e, -0.5)'$ is the point of the set of periodic regimes equidistant (for the projective distance) from $\pi(O_1^2)$ and $\pi(O_2^2)$. A periodic regime consists of two points equidistant from point $\pi(o)$.
The interpretation of the domains of attraction is the same as previously.

We have now to consider a special case of scs1-cyc2 matrices. It is a matrix whose critical graph contains two circuits of length 2. Let P be such a matrix. The critical circuits of the graph of P are $(1,2)$ and$(2,3)$. From the point of view of its behaviour, this matrix is not very different from a scs1-cyc2 matrix with only one critical circuit of length 2. In Figure 3.11, we compare P with a matrix having the same set of periodic regimes but only one critical circuit of length 2, matrix Q.

$$P = \begin{pmatrix} . & e & . \\ e & . & e \\ . & e & . \end{pmatrix}, \ Q = \begin{pmatrix} . & e & . \\ e & . & . \\ . & e & . \end{pmatrix}, \ (.) = -1 \, .$$

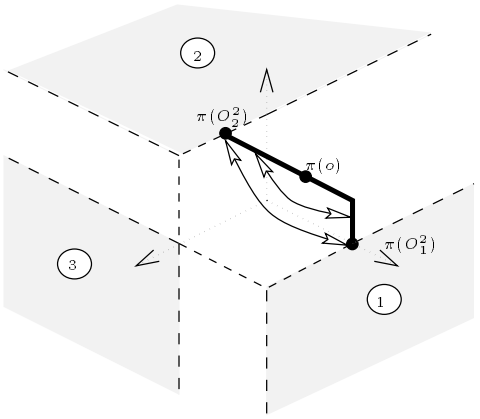


Figure 3.10: scs1-cyc2, a non-symmetrical example, set of periodic regimes of O.

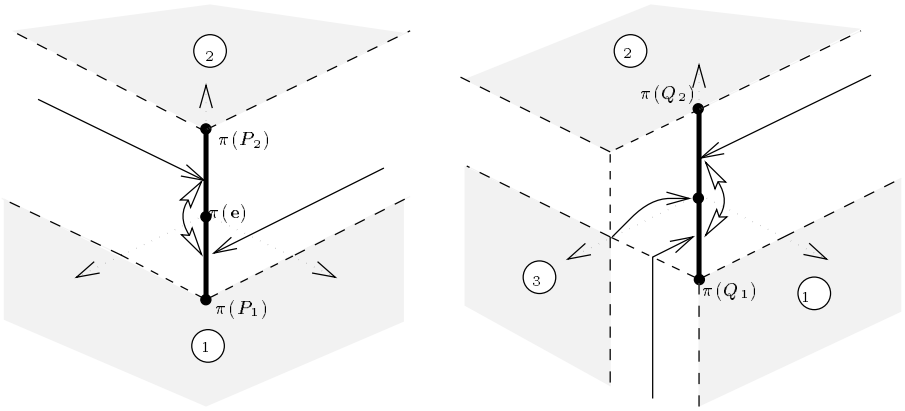


Figure 3.11: scs1-cyc2, a special case with two critical circuits of length 2, set of periodic regimes of P and Q.

For matrix P, the limit regime is $(\pi(P_1), \pi(P_2))$ for an initial condition in zone 1 or 2. The unique eigenvector, $\pi(\mathbf{e}) = \pi(e, e, e)'$ has a domain of attraction restricted to itself. For matrix Q, the limit regime is $(\pi(Q_1), \pi(Q_2))$ for an initial condition in zone 1 or 2. For an initial condition in zone 3, it is the eigenvector $\pi(\mathbf{e})$.

Remark All matrices having the same critical graph as P have exactly the same spectral behaviour as P. More precisely, let us consider a matrix $\tilde{P}$ obtained by modifying the non-critical terms. Its stationary version is denoted $\tilde{P}^N$ (Definition 3.4.8).

$$\tilde{P} = \begin{pmatrix} \alpha_1 & e & \alpha_2 \\ e & \alpha_3 & e \\ \alpha_4 & e & \alpha_5 \end{pmatrix}, \ \alpha_i < e, \ \tilde{P}^N = \begin{pmatrix} \alpha & e & \alpha \\ e & \alpha & e \\ \alpha & e & \alpha \end{pmatrix}, \ \alpha = \max_{i=1\ldots5} \alpha_i \, .$$

According to Stage 6 of the algorithm, the figure corresponding to $\tilde{P}$ is obtained from the one of P by an homothetic transformation of center $\pi(\mathbf{e}) = \pi(e, e, e)'$ and ratio $|\alpha|$.

Remark In this section, matrices I to O are such that the critical columns are 1 and 2. If we consider a matrix which is scs2-cyc1 or scs1-cyc2 but with different critical columns, one can get back to the previous cases by a permutation of the coordinates. It means that the corresponding picture can be obtained from the ones of this Section by performing a rotation of center **e**.
The same kind of remark applies to matrices which are scs1-cyc2 and have 2 critical circuits. All figures corresponding to such matrices can be obtained by a rotation of Figure 3.11, matrix P. The same kind of remark is also valid in the forthcoming case of scs2-cyc2 matrices

3.5.4 Scs2-Cyc2

The basic example of such a matrix is :

$$R = \begin{pmatrix} . & e & . \\ e & . & . \\ . & . & e \end{pmatrix}, \ (.) = -1 \, .$$

If a matrix M is scs2-cyc2 then the matrix M^2 is scs3-cyc1. To find the set of eigenelements of a scs2-cyc2 matrix M, one has to determine the set of eigenvectors of the scs3-cyc1 matrix M^2 (see Section 3.5.2).

Let us represent graphically eigenvectors, periodic regimes of period 2 and domains of attraction of matrix R in Figure 3.12.

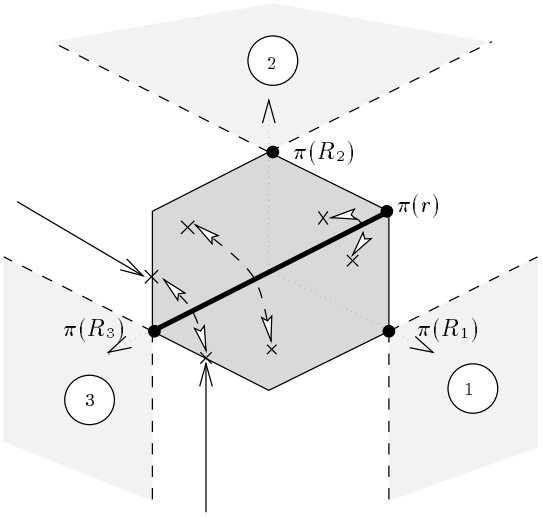


Figure 3.12: scs2-cyc2, set of eigenelements of R.

There is a symmetry axis for the whole figure (matrix R is unchanged by a permutation of the first two coordinates). There are two extremal eigenvectors, $\pi(R_3)$ and $\pi(r) = \pi(e, e, -1)'$. The set of eigenvectors (the interval $[\pi(R_3), \pi(r)]$) splits the set of periodic regimes in two equal parts. The two points of a periodic regime of period 2 are symmetric with respect to the set of eigenvectors. The analysis of domains of attraction is analog to the one of the scs3-cyc1 case. If the initial condition belongs to the zones 1 or 2 (resp. 3), the limit value is the periodic regime $\{\pi(R_1), \pi(R_2)\}$ (resp. the eigenvector $\pi(R_3)$). If the initial condition belongs to one of the three

white strips, the limit regime is a periodic regime of period 2, corresponding to the "nearest" point on the hexagon and its symmetrical point.

We have now to analyze what happens if we modify non-critical terms. The cases we have already considered are enough to understand what is going to happen. We will represent two characteristic examples in Figure 3.13, corresponding to matrices S and T.

$$S = \begin{pmatrix} a & e & . \\ e & . & . \\ b & . & e \end{pmatrix}, \; T = \begin{pmatrix} . & e & c \\ e & . & d \\ . & . & e \end{pmatrix}, \; (.) = -1 \; .$$

The reals a, b, c and d must satisfy the following constraints (in order for our matrices S and T to be scs2-cyc2) :

$$-1 \leqslant a < e, \; -1 \leqslant b < 1, \; -1 \leqslant c < 1, \; -1 \leqslant d < 1 \; .$$

The stationary versions of the matrices are :

$$S^3 = \begin{pmatrix} a & e & . \\ e & a & . \\ b & b & e \end{pmatrix}, \; T^3 = \begin{pmatrix} . & e & c \oplus d \\ e & . & c \oplus d \\ . & . & e \end{pmatrix}, \; (.) = -1 \; .$$

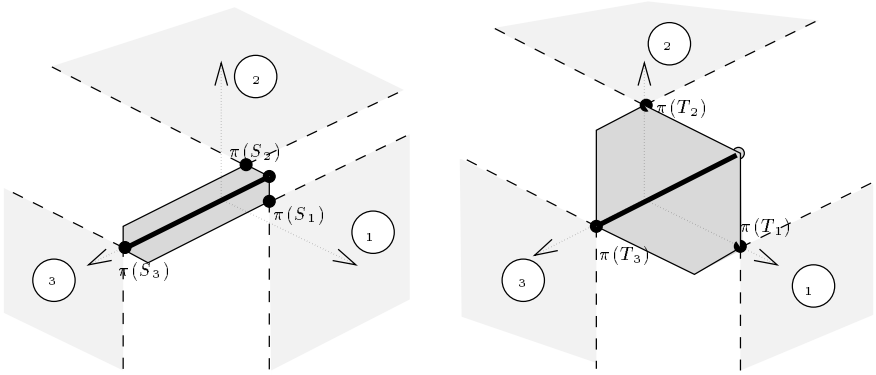


Figure 3.13: scs2-cyc2, sets of eigenelements of S and T.

The graphical representations of Figure 3.13 correspond to $a = -0.2, b = -0.5, c = -0.5$ and $d = -0.9$.

3.5.5 Transient regimes

We will now take a closer look at transient regimes of matrices. The matrices we have been considering so far were chosen in order to be stationary or at least to have a very short transient regime. To emphasize the transient behaviour, we will, on the other hand, consider matrices with long transient regimes. The length of the transient regime is closely related to the "second eigenvalue" of the matrix, i.e the second largest circuit weight (see [44]).

First of all, one has to remark that a matrix can have an arbitrarily long transient regime. Let us take an example.

$$U = \begin{pmatrix} e & -1 \\ -1 & -\eta \end{pmatrix},\ 0 < \eta \ll 1,\ U^2 = \begin{pmatrix} e & -1 \\ -1 & -2 \times \eta \end{pmatrix},$$

$$U^n = \begin{pmatrix} e & -1 \\ -1 & -n \times \eta \end{pmatrix},\ n < [\frac{2}{\eta}] + 1,\ \ U^n = \begin{pmatrix} e & -1 \\ -1 & -2 \end{pmatrix},\ n \geqslant [\frac{2}{\eta}] + 1\,.$$

The length of the transient regime is thus $[\frac{2}{\eta}]$. The matrix U is scs1-cyc1, its unique eigenvector is $\pi(u) = \pi(e, -1)'$. As we have seen previously, it implies that $\lim_n \pi(U^n v) = \pi(u)$, $\forall v \in \mathbb{R}^k$. Let us consider the initial condition $v = (e, 3)'$. We have $\pi(Uv) = \pi(e, 1 - \eta)'$, $\pi(U^2 u) = \pi(e, 1 - 2 \times \eta), \ldots$.

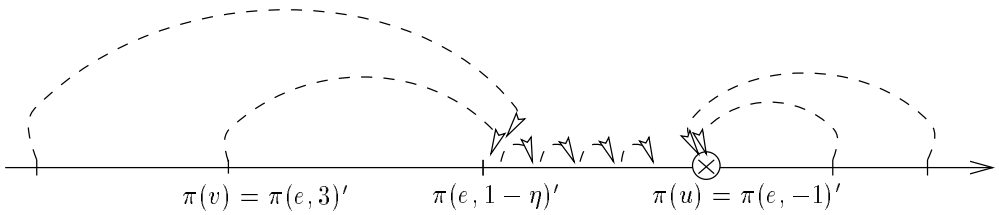


Figure 3.14: scs1-cyc1, dimension 2, transient regime of U.

We have represented on Figure 3.14, the sequence $\{\pi(U^n v)\}$ in the projective space $\mathbb{PR}^2$. We have also represented the same sequence for three other initial conditions. We are now going to present analog figures corresponding to matrices of size 3.

First of all, we consider the example of scs1-cyc1 matrices.

$$V = \begin{pmatrix} e & . & . \\ . & . & -\eta \\ . & -\eta & . \end{pmatrix},\ W = \begin{pmatrix} e & . & . \\ . & -2 & -\eta \\ . & -\eta & -2 \end{pmatrix},\ 0 < \eta \ll 1,\ (.) = -1\,.$$

Both matrices have the same stationary version :

$$\lim_k V^k = \lim_k W^k = \begin{pmatrix} e & . & . \\ . & -2 & -2 \\ . & -2 & -2 \end{pmatrix},\ (.) = -1\,.$$

We have represented the transient behaviours of matrices V and W in Figure 3.15. Matrix V is obtained by a small perturbation of matrix R (Section 3.5.4, Figure 3.12). The transient behaviour reflects it, as the figure we obtain is very close to Figure 3.12. As a comparison, we have also represented the matrix W whose behaviour is asymptotically identical.

Let us comment on the figure corresponding to V a little further. The three points $\pi(V_1), \pi(V_2)$ and $\pi(V_3)$ are the projections of the columns of matrix V. If the initial condition is in zone 1,

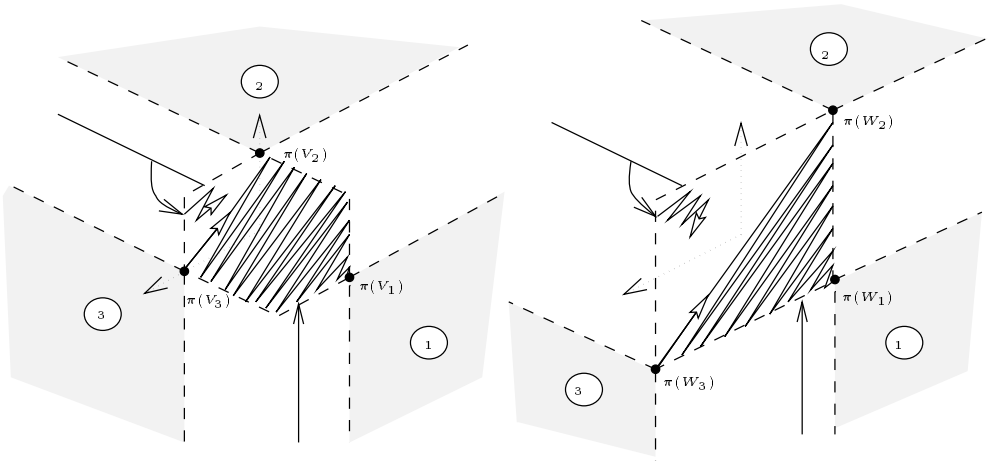


Figure 3.15: Dimension 3, scs2-cyc2, transient regimes of V and W.

there is convergence in one step to $\pi(V_1) = \pi(e, -1, -1)'$, the unique eigenvector. If the initial condition is in zone 2 (resp. 3), we have $\pi(Vx_0) = \pi(V_3) = \pi(-1, -1, -\eta)'$ (resp. $\pi(Vx_0) = \pi(V_2) = \pi(-1, -\eta, -1)'$). We have represented the whole sequence $\{\pi(V^n x_0)\}$ for an initial condition $\pi(x_0) = \pi(V_3)$. For an initial condition in one of the three white strips, for example let us consider u_0 (or u_0'), then $\pi(Vu_0)$ (or $\pi(Vu_0')$) is the point pointed by the arrow in the picture (it is the symmetric of the "nearest" point on the set $Im(S)$). For initial condition u_0', we have also drawn the beginning of the sequence $\{\pi(V^n u_0')\}$.

For matrix W, the set of periodic regimes is the same one as V. But the domains of attraction are quite modified. It emphasizes the possible influence of transient regimes, especially in stochastic models. Here we have drawn the sequences $\{\pi(V^n u_0)\}$ (or part of them) for several different initial conditions. One of them is in zone 2, another one in the white strip between zones 2 and 3 and the last one is on the symmetry axis.

We consider now the transient regime of a scs2-cyc1 matrix.

$$Z = \begin{pmatrix} e & . & . \\ . & e & . \\ . & . & -\eta \end{pmatrix}, \; 0 < \eta \ll 1, \; (.) = -1 \, .$$

The stationary regime of Z is matrix J, the basic scs2-cyc1 matrix (Section 3.5.3). We can also view Z as a small perturbation of matrix A the basic scs3-cyc1 matrix (Section 3.5.2. Figure 3.16 reflects these remarks.
The extremal eigenvectors are $\pi(Z_1)$ and $\pi(Z_2)$, the first two columns of Z. The point $\pi(z)$ is $\pi(e, e, 1-\eta)'$. If the initial condition is in zone 1 (resp. 2) we have convergence to $\pi(Z_1)$ (resp. $\pi(Z_2)$) in one step. The hexagon represented in dotted lines is $Im(Z) = Z(\mathbb{R}^3)$. For an initial condition $\pi(u)$ in one of the white strips, $\pi(Zu)$ is the closest (for the projective distance) point on the hexagon $Im(Z)$. Then there is convergence of $\pi(Z^n u)$ to the eigenvector of Z which is the closest to $\pi(Zu)$ (it is not necessary the closest to $\pi(u)$!) We have represented, in Figure 3.16, the whole sequence $\{\pi(Z^n x_0)\}$ for several different initial conditions.

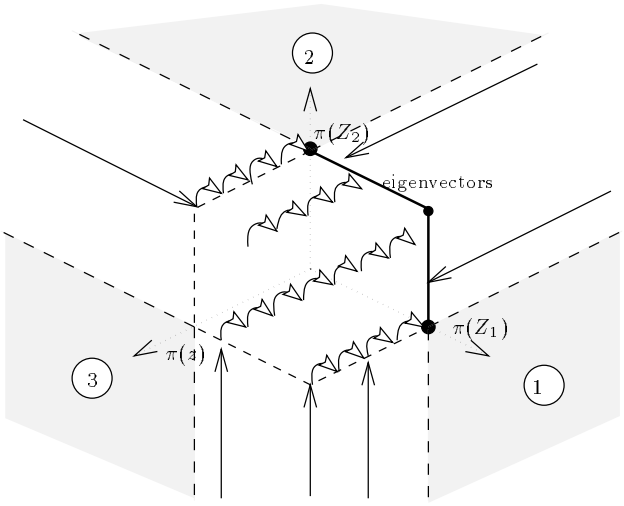


Figure 3.16: Dimension 3, scs2-cyc1, transient regime of Z.

3.6 Application to the Manufacturing Model

We consider the manufacturing model of Section 3.3, Figure 3.1. Matrix M in (3.4) is either scs1-cyc2 or scs1-cyc1, depending on the values of α, β_i and γ_i. As an example, let $\alpha = 2$, $\beta_1 = 2$, $\beta_2 = 1$, $\gamma_1 = 2$ and $\gamma_2 = 0$. We obtain :

$$x(n+1) = M \otimes x(n),\ M = \begin{pmatrix} \varepsilon & \varepsilon & 2 \\ \varepsilon & \varepsilon & 1 \\ 4 & 2 & 2 \end{pmatrix} .$$

The eigenvalue of the matrix is :

$$\lambda = \frac{M_{13} + M_{31}}{2} = 3 .$$

The circuit $(1,3)$ is the unique critical circuit. Hence matrix M is scs1-cyc2. The unique (in the projective space) eigenvector is $\pi(u_0) = \pi(1, e, 2)'$. The set of periodic regimes is $\{\pi(u_\eta), \pi(v_\eta)\} = \{\pi(1+\eta, \eta, 2), \pi(1, e, 2+\eta)\}, \eta \in [e, 1]$. We have represented the set of eigenelements of matrix M in Figure 3.17.
Let us consider a specific periodic regime $\{u_\eta, v_\eta\}$. We suppose that the system is in this regime, i.e. $\pi(x(2n)) = \pi(u_\eta)$ and $\pi(x(2n+1)) = \pi(v_\eta)$. We recall the definition of the idle time as given in Section 3.3, $\delta(n) = x_3(n) - x_3(n-1) - \alpha$. We have :

$$\delta(2n) = 1 + (u_\eta)_3 - (v_\eta)_3,\ \ \delta(2n+1) = 1 + (v_\eta)_3 - (u_\eta)_3 .$$

We see that the idle time of the assembly line depends on the periodic regime of the system.
In many practical cases, it is interesting to have a stationary regime such that the idle time is constant. It implies that the stationary regime has to be the eigenvector u_0. We also want to have a control which consists in choosing the initial condition. We conclude that we have to

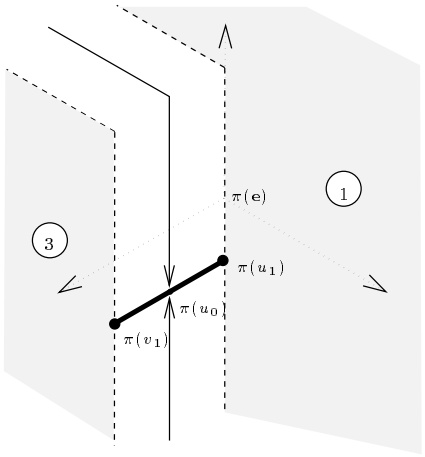


Figure 3.17: scs1-cyc2, manufacturing model, set of periodic regimes of M.

choose the initial condition in the domain of attraction of u_0. This domain can be observed on Figure 3.17. It corresponds to the vectors :

$$\{\pi(1,x,2)',\ x \in]-\infty,3]\} \cup \{\pi(x,2,1)',\ x \in]-\infty,e]\}\,. \tag{3.7}$$

Suppose now that we want to minimize the minimal idle time. We still want to have a control which consists in choosing the initial condition. The minimal idle time is :

$$\underline{\delta} = \min \delta(2n), \delta(2n+1) = \min(1-\eta, 1+\eta) = 1-\eta\,.$$

So we minimize $\underline{\delta}$ for $\eta = 1$, i.e. for the periodic regime $\{u_1, v_1\}$. We have to choose the initial condition in the domain of attraction of $\{u_1, v_1\}$. This domain is the union of zones 1 and 3 on Figure 3.17.

The graphical representation has enabled us to illustrate a control problem in a manufacturing model. As simple as the problem is, we see that the sets of solutions (for example the one given in (3.7)) are far from being simple or intuitive. It enables one to get an idea of the potential complexity of this kind of problems.

3.7 A Projectively Infinite Semigroup of Matrices

We consider a finite number of matrices $A_1, \ldots, A_p \in \mathbb{R}_{max}^{k\times k}$. We denote respectively by $< A_1, \ldots, A_p >$ and $\pi< A_1, \ldots, A_p >$ the semigroup generated by $A_1, \ldots, A_p$ and its projection.

$$< A_1, \ldots, A_p >= \{(A_{u_N} \cdots A_{u_2} A_{u_1})\,,\ u_1, \ldots, u_N \in \{1, \ldots, p\}, N \text{ finite}\}\,,$$

$$\pi< A_1, \ldots, A_p >= \{\pi(A_{u_N} \cdots A_{u_2} A_{u_1})\,,\ u_1, \ldots, u_N \in \{1, \ldots, p\}, N \text{ finite}\}\ ,$$

where π is here the canonical projection of $\mathbb{R}_{max}^{k\times k}$ into $\mathbb{PR}_{max}^{k\times k}$. The problem we are interested in is the finiteness of $\pi< A_1, \ldots, A_p >$. It is in fact a version in the (max,+) algebra of the classical Burnside problem (see Gaubert [70]).

Let us consider the projective semigroup generated by a single irreducible matrix $\pi < A >= \{\pi(A^n), n \in \mathbb{N}\}$. Theorem 3.4.7 tells us that $\pi< A >$ is finite.

Remark It is the finiteness of the projective semigroup and not the finiteness of the semigroup which is interesting. Indeed any irreducible matrix A with an eigenvalue different from e is such that $< A >$ is infinite.

A slightly stronger version of next theorem was proved in [70], see also Proposition 6.5.4.

Theorem 3.7.1. *Let $A_1, \ldots, A_p \in \mathbb{Q}_{max}^{k\times k}$. We assume that :*

$$\forall u \in \{1, \ldots, p\}, \forall (i,j),\ (A_u)_{ij} > \varepsilon\ .$$

Then the projective semigroup $\pi< A_1, \ldots, A_p >$ is finite.

This theorem can not be extended to the case of matrices with non rational entries. We are going to propose a counter-example.
We consider the semigroup generated by the matrices :

$$A_1 = \begin{pmatrix} -\eta_1 & . & . \\ . & e & . \\ . & . & e \end{pmatrix},\ A_2 = \begin{pmatrix} e & . & . \\ . & -\eta_2 & . \\ . & . & e \end{pmatrix},\ A_3 = \begin{pmatrix} e & . & . \\ . & e & . \\ . & . & -\eta_3 \end{pmatrix},$$

where $(.) = -1$, $0 < \eta_i \ll 1$ and $\eta_i \notin \mathbb{Q}$. We suppose also that $\eta_i/\eta_j \notin \mathbb{Q}$, $i, j \in \{1,2,3\}$, $i \neq j$. An easy way to show that the semigroup $\pi< A_1, A_2, A_3 >$ is infinite is to consider the initial condition $\mathbf{e} = (e, e, e)'$ and to prove that $\Pi = \pi(< A_1, A_2, A_3 > \mathbf{e}) = \{\pi(M\mathbf{e}),\ M \in< A_1, A_2, A_3 >\}$ is infinite. We obtain a nice illustration of the phenomenon with the help of the graphical representation in the projective space.

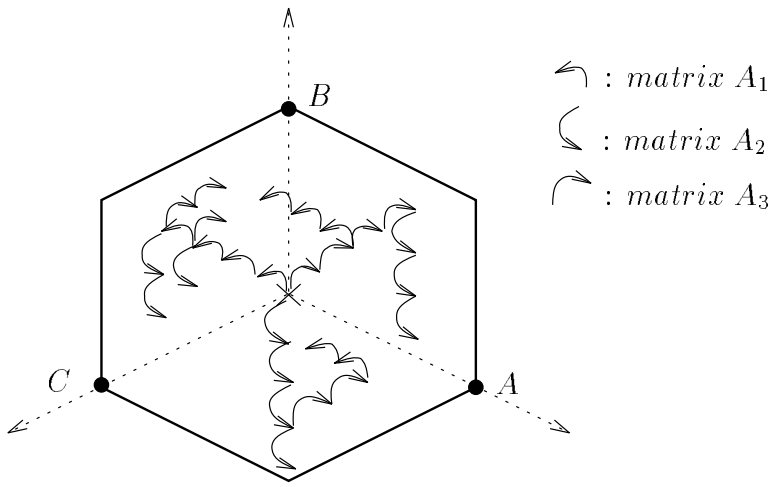


Figure 3.18: A finitely generated but projectively infinite semigroup of matrices.

The extremal eigenvectors of A_1, A_2 and A_3 are respectively $(\pi(e_2), \pi(e_3))$, $(\pi(e_1), \pi(e_3))$ and $(\pi(e_1), \pi(e_2))$. Figure 3.18 shows the effect of applying matrices A_1, A_2 and A_3 to a vector. Figure 3.18 is analog to Figure 3.16 but with three different transient regimes interacting.

For a point $u = (u_1, u_2, u_3)'$ such that $d(u, \mathbf{e}) < 1 - \sup_{i=1,2,3} \eta_i$, where d is the projective distance (Definition 3.4.10), we have :

$$A_1 u = \begin{pmatrix} u_1 - \eta_1 \\ u_2 \\ u_3 \end{pmatrix}, \; A_2 u = \begin{pmatrix} u_1 \\ u_2 - \eta_2 \\ u_3 \end{pmatrix}, \; A_3 u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 - \eta_3 \end{pmatrix} .$$

It is easy to prove that Π is dense in the hexagon $\mathcal{H}$ delimited by $\pi(e_1), \pi(e_1)$ and $\pi(e_3)$. In fact let us consider three integers N_1, N_2 and N_3 such that :

$$\sup_{i=1,2,3} (N_i \times \eta_i) - \inf_{i=1,2,3} (N_i \times \eta_i) < 1 .$$

Then it is quite obvious that there exists a matrix $M \in < A, B, C >$, $M = A_{u_N} \otimes \cdots \otimes A_{u_1}$ verifying : with $N = N_1 + N_2 + N_3$ where :

$$N = N_1 + N_2 + N_3, \quad N_i = \#\{n \mid A_{u_n} = A_i\}, \; i = 1, 2, 3 ,$$

$$M\mathbf{e} = \begin{pmatrix} -N_1 \times \eta_1 \\ -N_2 \times \eta_2 \\ -N_3 \times \eta_3 \end{pmatrix} .$$

In fact it is easy to understand, watching Figure 3.18, that we will obtain this formula for $M\mathbf{e}$ iff $\forall n \in \{1, \ldots, N\}$, $\pi(A_{u_n} \otimes \cdots \otimes A_{u_1}\mathbf{e})$ belongs to the interior of the hexagon $\mathcal{H}$.

Let us consider an arbitrary point $\pi(v)$ in the interior of the hexagon $\mathcal{H}$. As η_1, η_2, η_3 are not co-rational, there exists a sequence of integers $N^{(n)}$ and a sequence of matrices $\{M^{(n)}, M^{(n)} \in < A_1, \ldots, A_p >\}$ with the following properties.

- The length of $M^{(n)}$ is $N^{(n)}$, i.e $M^{(n)} = A^{(n)}_{u_{N^{(n)}}} \otimes \cdots \otimes A^{(n)}_{u_1}$.
- $N_i^{(n)} = \#\{l \mid A^{(n)}_{u_l} = A_i\}, \; i = 1, 2, 3,$

$$\pi(M^{(n)}\mathbf{e}) = \pi \begin{pmatrix} -N_1^{(n)} \times \eta_1 \\ -N_2^{(n)} \times \eta_2 \\ -N_3^{(n)} \times \eta_3 \end{pmatrix} \stackrel{n \to \infty}{\longrightarrow} \pi(v) .$$

Remark If we consider another initial condition $u \neq \mathbf{e}$, we will in general obtain a set of reachable points $\pi(< A_1, \ldots, A_p > u)$ dense in the hexagon $\mathcal{H}$ and whose intersection with Π is empty. Let us now consider a Markov chain $x(n, x_0)$ whose transition probabilities $p(.,.)$ verify

$$\forall v \in \mathbb{R}^3_{max}, \; p(\pi(v), \pi(A_i v)) = p_i, \; i = 1, 2, 3, \; p_i > 0, \; p_1 + p_2 + p_3 = 1 .$$

We take $\pi(\mathbf{e})$ as our initial condition. Then Π is a set of transient states for the Markov chain. Between the first and and the second hitting of the border of the hexagon $\mathcal{H}$, the Markov chain evolves on a set of transient states dense in the interior of $\mathcal{H}$ and whose intersection with Π is empty. It is however possible to show that the chain is positive recurrent. Points $\pi(e_1), \pi(e_2)$ or $\pi(e_3)$ can be used as regenerative points (for example $\pi(A_2^{n''} A_3^{n'} A_2^n u) = \pi(e_1)$, $\forall u \in \mathbb{R}^k$, when n, n' and n'' are sufficiently large).

3.8 Conclusion

The main contribution of this paper is the graphical characterization of the domains of attraction of the eigenelements of a matrix. The drawback of the approach is that it considers only matrices of size 3. However, it should be noted that for a general matrix, if the critical circuit is of size 3 or less, then the proposed approach applies. Moreover, the study of these 3x3 matrices provides a good intuition of the general case. For example, in any dimension, the domains of attraction will be "polyhedrons" and the sets of eigenelements, compact polyhedrons.

A C program has been written by Bruno Gaujal, which implements the algorithm of Section 3.5. Given a matrix of dimension 3, this program provides the graphical representation of eigenvectors, periodic regimes and domains of attraction (as in Figures 3.4 to 3.13). If you are interested in obtaining this program, send a request to gaujal@@sophia.inria.fr or mairesse@@sophia.inria.fr.

Chapitre 4

Application in Cyclic Scheduling

Application à l'Ordonnancement Cyclique

On propose une application de la représentation graphique présentée au chapitre précédent. Il s'agit d'illustrer des problèmes propres à l'ordonnancement cyclique dans les systèmes de production.

Ce chapitre a bénéficié de nombreuse discussions avec Zhen Liu et Lucian Finta.

4.1 Introduction

The scheduling problems that we are going to consider appear in manufacturing systems. We have sets of jobs (or tasks) and machines. In general, the scheduling problem is to map in the most efficient way the jobs on the machines. Our problem is a little bit different. We assume that the mapping is given and we want to determine, in an optimal way, the initial delays between jobs.

The basic model can be represented in the form of an oriented graph G with weights (in $\mathbb{R}^+$) on the nodes and the arcs. A node corresponds to some tasks and its weight corresponds to the execution time of that task. An arc corresponds to a precedence relation between two tasks and its weight corresponds to the communication time between the tasks. Let us assume that there is an arc from task i to task j with weight δ. It means that the execution of task j can start only δ units of time after the completion of task i. Note that weights on nodes and arcs are called execution and communication times respectively

One must distinguish between two different classes of problem.

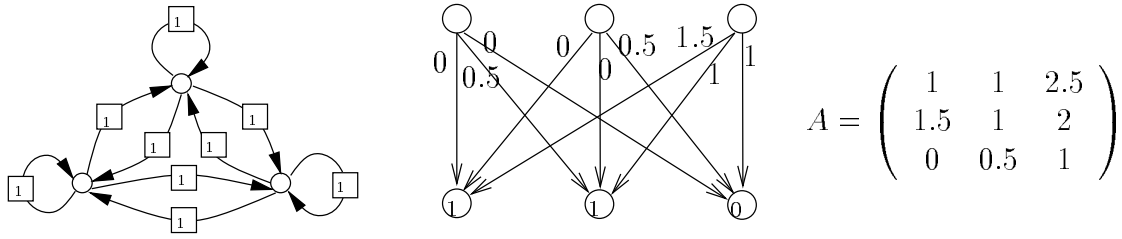
1. Classical scheduling problem. In this case, there is a finite number of tasks to be executed.
2. Cyclic scheduling problem. There is a finite number of *generic tasks*, but each generic task has to be processed an infinite number of times. This problem is modelled with a graph as defined above. However, it is necessary to add a *delay* (with value in $\mathbb{N}$) on arcs. Let $(i,n), n \in \mathbb{N}$, denote the n-th execution of task i. If there is an arc from task i to task j with delay l, it means that there are precedences from the tasks $(i,n), \forall n$, to the tasks $(j,n+l), \forall n$. This graph is precisely the reduced graph presented in Chapter 2 §2.2.2.

In the following, we consider the cyclic scheduling problem. It appears naturally to model mass production in manufacturing systems.

There are different types of problems which can be considered within the framework of cyclic scheduling. For a complete review on the subject, see Hanen and Munier [84]. We restrict our attention to the so-called *basic cyclic scheduling problem*. It is assumed that the mapping of the tasks on the machines is completely defined. The optimal schedule is the earliest schedule, also called the "as soon as possible" schedule. Each task is executed as soon as all the preceding tasks have been executed.

We have described above the cyclic scheduling problem as a weighted graph. As detailed in Section §2.7, such a graph is also equivalent to an event graph. The as soon as possible evolution can be represented by a (max,+) linear system, see §1.3 and §2.7. Practically, it is done by constructing an equivalent graph where all the delays on arcs are equal to 1. This graph might have more nodes, i.e. more generic tasks, than the original one. When all the delays are equal to 1, we define a (max,+) matrix A. The coordinate A_{ij} corresponds to the sum of the execution time of task i and the communication time from task j to task i.
As an example, we have represented in Figure 4.1 the reduced graph, (a part of) the dependence graph and the (max,+) matrix of a given system. The graphical conventions are the ones of Chapter 2. On the reduced graph, we have represented only the delays. Execution and communication times are represented on the dependence graph.

Remark 4.1.1. It might appear unnatural to have execution and/or communication times which are equal to 0, as in Figure 4.1. Such a matrix has been chosen for the sake of simplicity. Exactly the same behaviour would be obtained for the matrix $A + \lambda$ where all the times have been increased by a same constant.



$$A = \begin{pmatrix} 1 & 1 & 2.5 \\ 1.5 & 1 & 2 \\ 0 & 0.5 & 1 \end{pmatrix}$$

Figure 4.1: Cyclic scheduling problem with 3 generic tasks.

We define the vectors of daters $x(n)$ where $x_i(n)$ corresponds to the instant of completion of the n-th occurrence of the generic task i. The vector x_0 is the initial condition. It corresponds to the initial delays of the generic tasks $(1,0),\ldots,(k,0)$. The as soon as possible schedule (given the initial condition x_0) corresponds to the (max,+) linear equation $x(n) = A^n \otimes x_0$.

From now on, we work with the (max,+) matrix $A \in \mathbb{R}_{max}^{k\times k}$ and we assume that this matrix is irreducible[1].

The main problem of interest in such a model is to choose the initial condition x_0 in an optimal way. From now on, we call "schedule" associated with an initial condition x_0 (in short : *schedule* x_0), the earliest execution pattern $\{x(n) = A^n x_0, n \in \mathbb{N}\}$.
There are two main criteria for the performance evaluation of schedules.

1. The cycle time. It is the average time of execution of the set of generic tasks, i.e $\lim_n \max_i x(n)_i/n$.

2. The latency. The latency is the time elapsed between the first completion of a generic task and the last completion of a generic task, i.e $\max_i x(n)_i - \min_i x(n)_i$.

The cycle time is the most important criterion. The latency is interesting, for example in manufacturing systems when the tasks (objects) have to be packed together at the end of the production line. Minimizing the latency will minimize the packing operation.

In our model, all schedules provide the same cycle time as a consequence of the spectral theory in the (max,+) algebra, see Chapter 3 Theorem 3.4.3. The cycle time is the maximal eigenvalue of matrix A. On the other hand, x_0 has a strong influence on the latency. When schedule x_0 is a periodic regime, the latency will also be periodic. In this case, we define the latency as being the average latency.

[1] In the reducible case, the latency, to be defined below, is often degenerate (equal to ∞). It is a non interesting case for our study.

Lemma 4.1.2. *Let $A \in \mathbb{R}_{max}^{k\times k}$ be an irreducible matrix. Let x_0 be an eigenelement. The latency of the schedule x_0 is $d(e, x_0)$ where $d(.,.)$ is the projective distance (see Chapter 3 Definition 3.4.9).*

The proof is a simple rephrasing of the definition of the latency. The problem we are going to address in the following is to choose x_0 in order to minimize the latency. This problem was considered by Lee [99]. In Parhi & Messerschmidt [116], they consider a specific question : is it possible to choose x_0 in order to have a latency equal to 0. In this Chapter, our goal is not to prove deep results. It is rather to illustrate the complexity of the phenomena involved and the interest of the graphical approach to understand them. However, we will prove some results for subclasses of (max,+) matrices in §4.4.

4.2 Periodicity of the Schedule

When deciding on a schedule, a first approach is to use the simplest one without computing the eigenelements of the matrix A. In such a case, the most natural choice is to consider the "as soon as possible" initial condition $e = (0, \ldots, 0)'$. By monotonicity, this choice yields the earliest execution of the daters $x_i(n)$, i.e.

$$(x(n, e) = A^n e) \leqslant (x(n, x_0) = A^n x_0)\,,\ \forall x_0\ :\ \forall i, (x_0)_i \geqslant 0\,. \tag{4.1}$$

On the other hand, such an initial condition might provide a transient regime (which can be arbitrary long, see §3.5.5) and a p-periodic schedule with $p > 1$. This last point is illustrated in Figure 4.2, where we propose the graphical representation of the matrix of Figure 4.1.

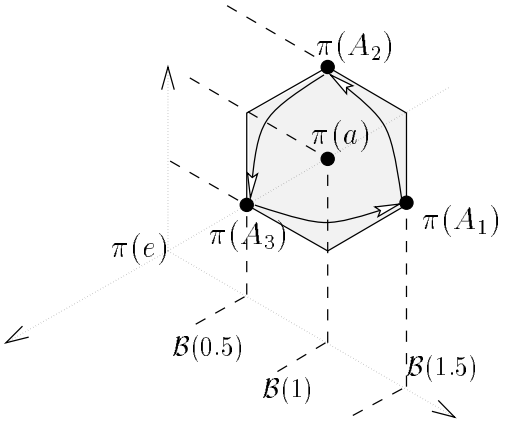


Figure 4.2: Graphical representation of matrix A.

Matrix A is scs1-cyc3. It has a unique eigenvector, $\pi(a) = \pi(1, 1, 0)$. We have represented the balls for the projective distance (see Chapter 3) of center $\pi(e)$ and respective radius 0.5, 1 and 1.5. Starting from the initial condition $e = (0, 0, 0)$, one obtains a 3-periodic schedule corresponding to the points $\pi(A_1), \pi(A_2)$ and $\pi(A_3)$. Starting from an initial condition $x_0 = (1, 1, 0)'$, we obtain a 1-periodic schedule. Let us represent the schedules e and $(1, 1, 0)$ using Gantt charts[2].

[2]Gantt charts are a common representation for scheduling problems. Here, it is simply a representation of the dependence graph incorporating the time.

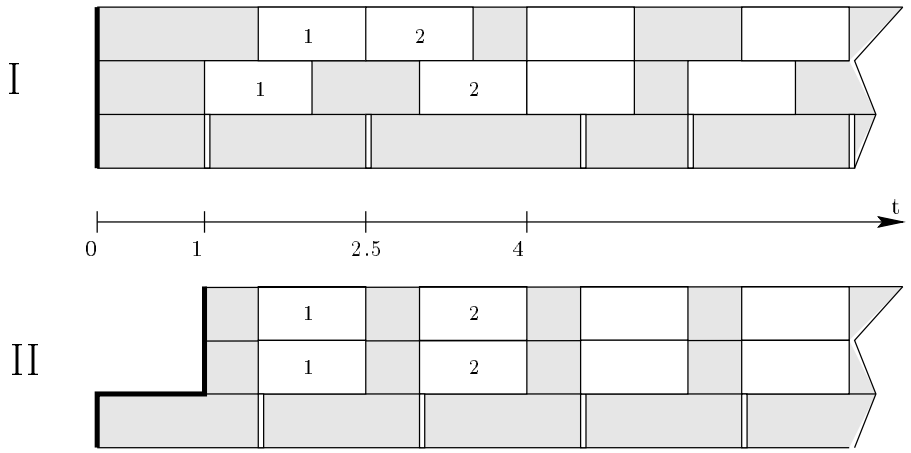


Figure 4.3: Chart I : schedule $e = (0,0,0)'$. Chart II : schedule $a = (1,1,0)'$.

We have materialized only the execution times and not the communication times on Figure 4.3. The thin bars correspond to the execution times equal to 0.

In some cases, it might be important to avoid p-periodic regimes. We are going to explain why using the language of parallel programs (but it is also relevant in manufacturing).

For parallel algorithms, there is an important distinction between schedules which are performed at run time (dynamic strategy) or at compile time (static strategy). Following the terminology of [131], we call them self-timed and fully-static schedules, respectively.

- Fully-static schedule. The instants of execution of the tasks are enforced by a finite state controller (automaton). The controller is designed before the execution, at compile time.
- Self-timed schedule. Each processor has to synchronize with other processors without an external control. This is done through the use of semaphore checks at run time.

For different reasons including the compared cost of controllers and semaphores, one might prefer one solution or the other. When the fully-static schedule is the best solution, it is important to minimize the number of states of the controller. In order to achieve this goal, it is necessary to have a schedule which is as simple as possible. The simplest schedules are the 1-periodic ones, they are associated with eigenvectors of the (max,+) matrix. For example, in Figure 4.3, the 1-periodic schedule of chart II is associated with the eigenvector $(1,1,0)$ of matrix A.

To summarize what precedes, here are the three criteria which we would like, optimally, to see verified by a schedule x_0 :

- P : Periodicity 1 $\Rightarrow$ x_0 eigenvector.
- L : Minimal latency $\Rightarrow$ x_0 is such that $d(A^n x_0, e)$ is minimal.
- S : as Soon as possible (Equation (4.1)) and Simplicity (no pre-computation) $\Rightarrow$ $x_0 = e = (0,\ldots,0)'$.

More precisely, for a fully-static schedule the important criteria are P and L (S is irrelevant as the pre-computation is necessary to build the controller). For a self-timed schedule, the important criteria are L and S.

4.3 Graphical Illustration

We are going to illustrate the fact that the criteria P, L and S are not always compatible. We start with a favorable example.

All the criteria can be satisfied

Let us consider :

$$B = \begin{pmatrix} 2 & \varepsilon & \varepsilon \\ \varepsilon & 0.5 & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{pmatrix} \otimes \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix} \otimes \begin{pmatrix} -2 & \varepsilon & \varepsilon \\ \varepsilon & -0.5 & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{pmatrix} = \begin{pmatrix} 3 & 3.5 & 4 \\ 0.5 & 3 & 2.5 \\ 0 & 1 & 3 \end{pmatrix} . \qquad (4.2)$$

Matrix B has been written, in Equation (4.2), under a form which emphasizes that B is obtained as a translation of a canonical scs3-cyc1 matrix, see §3.5.2.

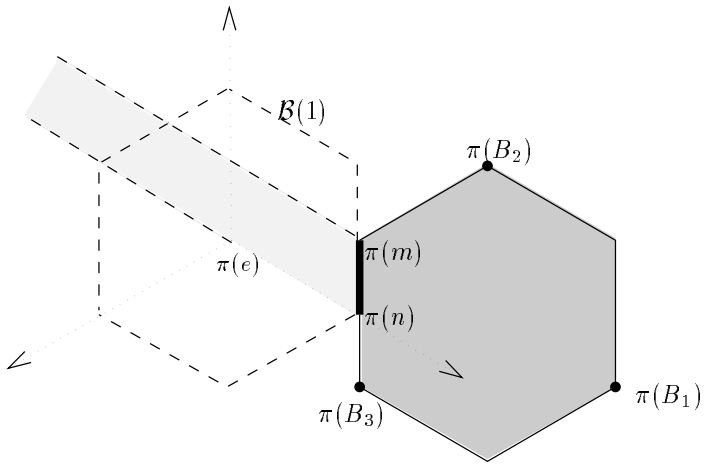


Figure 4.4: Criteria P, L and S are satisfied at the same time.

The graphical representation of matrix B is proposed in Figure 4.4. We have only eigenvectors, hence property P is always verified. We have represented $\mathcal{B}(1)$, the smallest ball with center $\pi(e)$ and intersecting the set of eigenelements. The intersection, the segment $[\pi(m), \pi(n)]$, verifies property L. We have represented in light gray the domain of attraction of $[\pi(m), \pi(n)]$. It contains $\pi(e)$. We conclude that criteria P, L and S are compatible.

The criteria cannot be satisfied together

In Figure 4.5, we consider:

$$C = \begin{pmatrix} 2 & \varepsilon & \varepsilon \\ \varepsilon & 1.5 & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{pmatrix} \otimes \begin{pmatrix} 2 & 2 & 3 \\ 3 & 2 & 2 \\ 2 & 3 & 2 \end{pmatrix} \otimes \begin{pmatrix} -2 & \varepsilon & \varepsilon \\ \varepsilon & -1.5 & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{pmatrix} = \begin{pmatrix} 2 & 2.5 & 5 \\ 2.5 & 2 & 3.5 \\ 0 & 1.5 & 2 \end{pmatrix}.$$

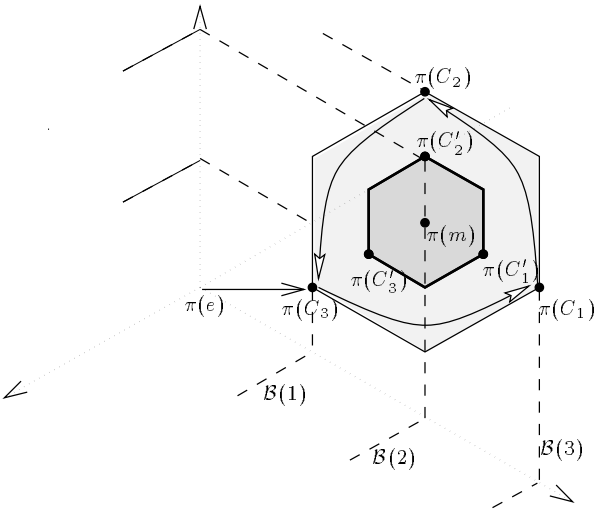


Figure 4.5: Criteria P and L are not compatible with S.

Matrix C is scs1-cyc3. The set of 3-periodic regimes is the ball delimited by the points $\pi(C_1), \pi(C_2)$ and $\pi(C_3)$. There is a unique eigenvector, $\pi(m) = \pi(2, 1.5, 0)'$, hence a unique 1-periodic schedule. The set of eigenelements having a minimal latency is the ball delimited by the points $\pi(C'_1), \pi(C'_2)$ and $\pi(C'_3)$. For example, the latency of $\pi(m)$ is $\delta = d(m, e) = 2$. The latency of the periodic regime (C'_1, C'_2, C'_3) is $\delta = 1/3 \times (d(C'_1, e) + d(C'_2, e) + d(C'_3, e)) = 1/3 \times (2.5 + 2 + 1.5) = 2$.
From the initial condition e, we obtain the periodic regime (C_1, C_2, C_3) whose latency is $\delta = 1/3 \times (3 + 3 + 1) = 7/3 > 2$.

In Figure 4.6, we consider:

$$D = \begin{pmatrix} -0.5 & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0.5 & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon \\ \varepsilon & \varepsilon & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1.5 & 1 \\ 2.5 & 1 & 1.5 \\ 1 & 0.5 & 0 \end{pmatrix}.$$

Matrix D is scs1-cyc2. The set of 2-periodic regimes is $[\pi(D_2), \pi(m)] \cup [\pi(m), \pi(D_1)]$. There is a unique eigenvector (1-periodic schedule) $\pi(m) = \pi(1, 1.5, 0)$. Its latency is $\delta = 1.5$. The set of eigenelements having a minimal latency is $[\pi(D_2), \pi(D'_2)] \cup [\pi(D'_1), \pi(D_1)]$. Their latency is $\delta = 1/2 \times (1 + 1.5) = 1.25$. The initial condition e belong to the domain of attraction of this set.

In order to further illustrate this example, we have represented, in Figure 4.7, the Gantt charts associated with the schedules e and $m = (1, 1.5, 0)'$.

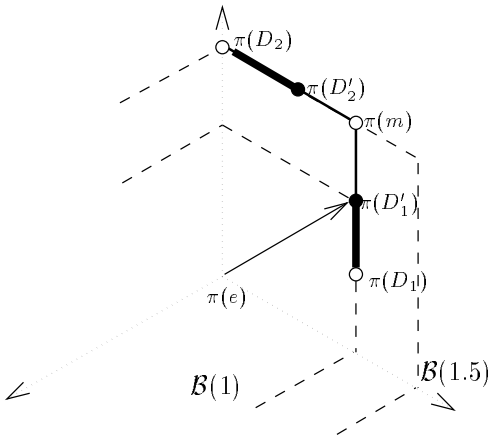


Figure 4.6: Criteria L and S are not compatible with P.

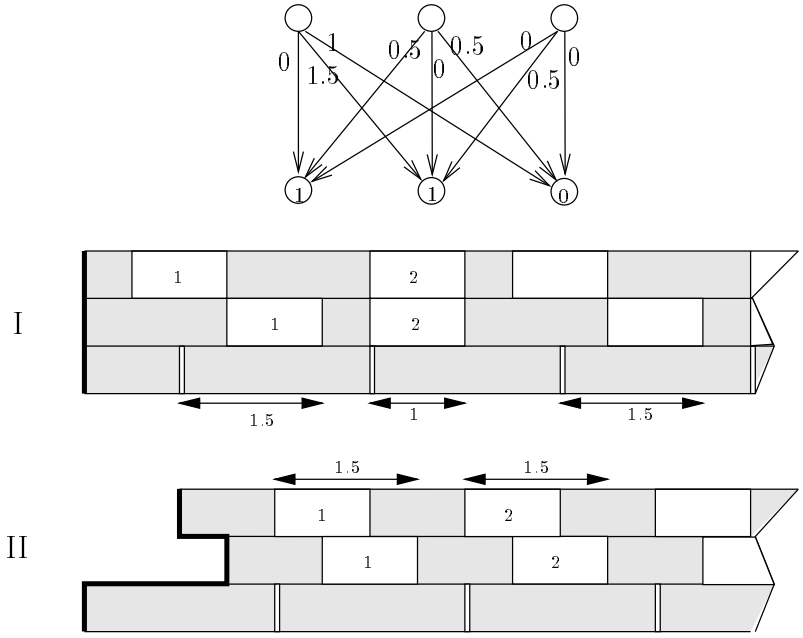


Figure 4.7: Chart I : schedule $e = (0, 0, 0)'$. Chart II : schedule $m = (1, 1.5, 0)'$.

In Figure 4.8, we consider

$$E = \begin{pmatrix} 0 & \varepsilon & \varepsilon \\ \varepsilon & 1 & \varepsilon \\ \varepsilon & \varepsilon & -1 \end{pmatrix} \otimes \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & \varepsilon & \varepsilon \\ \varepsilon & -1 & \varepsilon \\ \varepsilon & \varepsilon & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 3 \\ 3 & 3 & 4 \\ 1 & 0 & 1 \end{pmatrix} .$$

Matrix E is scs2-cyc1. The set of eigenvectors is $[\pi(E_2), \pi(m)] \cup [\pi(m), \pi(E_1)]$. The vector $e = (0, \ldots, 0)'$ belongs to the domain of attraction of $\pi(m)$. The latency of $\pi(m)$ is $\delta = 3$. The eigenvector $\pi(E_1)$ has a minimal latency $\delta = 2$.

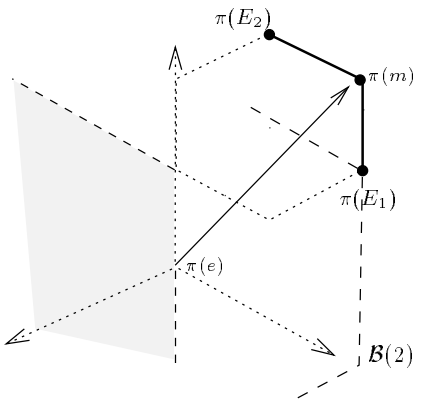


Figure 4.8: Criterion S is not compatible with L.

4.4 Quantitative Results

Here are some results for subclasses of (max,+) matrices.

Proposition 4.4.1. *Let $A \in \mathbb{R}_{max}^{k\times k}$ be an irreducible matrix. We suppose that all the nodes belong to the critical graph of A. Let Π be the set of eigenelements of A, i.e. $\Pi = \lim_n \mathrm{Im}(A^n) = \lim_n\{u \ : \ \exists v, A^k \otimes v = u\}$. For all $u \in \mathbb{R}^k$, the limit of $\pi(A^k \otimes u)$ is an eigenelement of A containing at least one point minimizing the distance $d(u,\Pi)$.*

For a proof, see Tronel [133].

Remark 4.4.2. Proposition 4.4.1 is not true for a general irreducible matrix A. For a counter-example, consider the matrix E represented in Figure 4.8 and the initial condition $e = (0,0,0)'$.

The example of matrix C, Figure 4.5, is interesting. The schedule e converges to the schedule associated with (C_1, C_2, C_3). Proposition 4.4.1 is verified as we have that $d(e, C_3) = 1$ which minimizes $d(e,\Pi)$. However the schedule (C_1, C_2, C_3) has not a minimal average latency (as $d(e, C_1) = d(e, C_2) = 3$).

Corollary 4.4.3. *We assume furthermore that A is scsk-cyc1. The schedule $e = (0,\ldots,0)'$ converges to a 1-periodic schedule having minimal latency.*

Proof. As a consequence of Theorem 3.4.7, we have that there exists a finite n such that $\pi(A^n e)$ is an eigenvector of A. We achieve by applying Proposition 4.4.1. □

Proposition 4.4.4. *We assume now that matrix A is scs1-cyck, i.e. has a unique critical circuit of length k. Let a be the unique eigenvector of matrix A. The schedule $\{x_0 = a\}$ has a minimal latency.*

Proof. We denote by $v = (v_1, v_2, \ldots, v_k)'$ a periodic regime of period k of A. For a scs1-cyck matrix, the eigenvector a is the "barycentre" of the points $(v_1, v_2, \ldots, v_k)'$. It means that

$$\begin{aligned} \pi(a) &= \pi\left(\frac{v_1 + v_2 + \cdots + v_k}{k}\right) \\ \Rightarrow \pi(k \times a) = \pi(a^{\otimes k}) &= \pi(v_1 + v_2 + \cdots + v_k)\ . \end{aligned}$$

This result is due to Braker and Olsder [29], see also §5.3. We deduce that we have :

$$\begin{aligned} k \times d(e,a) &= d(e,a^{\otimes k}) = d(e, v_1 + v_2 + \cdots + v_k) \\ &\leqslant d(e,v_1) + d(e,v_2) + \cdots + d(e,v_k) \\ \Longrightarrow d(e,a) &\leqslant \frac{1}{k} \times (d(e,v_1) + d(e,v_2) + \cdots + d(e,v_k)) \end{aligned}$$

It proves precisely that a has a minimal latency. □

An illustration of Proposition 4.4.4 is provided by matrix C, Figure 4.5. It is known, see [29] or §5.3, that the equality $\pi(a) = \pi(v_1 + \cdots v_l/l)$ does not always hold when the matrix is scs1-cyc$l$, $l < k$. In this case, we cannot conclude. A counter-example is provided by matrix D, Figure 4.6, where the unique eigenvector has a non minimal latency.

Chapitre 5

Algorithms

Algorithmes

On propose une méthode permettant de calculer tous les vecteurs propres **et** tous les régimes périodiques d'une matrice (max,+). Les algorithmes proposés sont classiques. Leur application pour l'obtention de la valeur propre et des vecteurs propres d'une matrice (max,+) aussi, voir par exemple [78] et [44]. La nouveauté (à notre connaissance) est leur utilisation pour obtenir l'ensemble des régimes périodiques. Cela nécessite une transformation préalable de la matrice. On considère également un algorithme alternatif introduit dans [29]. On analyse cet algorithme à l'aide de l'approche graphique du chapitre 3.

We consider an irreducible matrix $A \in \mathbb{R}_{max}^{k\times k}$. The aim is to compute its eigenvalue and all its eigenelements, i.e. all the eigenvectors and periodic regimes, of A. We propose an algorithm with overall complexity in $O(k^3\sqrt{k\log k})$.

5.1 Algorithm

We consider $A \in \mathbb{R}_{max}^{k\times k}$. We propose the following algorithm.

1. Determine if A is irreducible.
2. Determine if A is aperiodic.
3. Compute the eigenvalue of A.
4. Compute all the eigenvectors of A.
5. Determine the spectral type of A.
6. Compute all the eigenelements of A.

Let us detail the different stages.

Stage 1 (Determine if A is irreducible).
We consider the boolean matrix associated with A. It is defined in the following way: $\tilde{A}_{ij} = \varepsilon$ if $A_{ij} = \varepsilon$, $\tilde{A}_{ij} = e$ if $A_{ij} > \varepsilon$. We consider the graph of this matrix (see Def. 3.4.1). We apply an algorithm of Tarjan, see [132]. It provides all the maximal strongly connected subgraphs (s.c.s) of an oriented graph. It is a "depth first search" algorithm. It requires the construction of a covering tree of the graph.

From now on, we assume that matrix A is irreducible i.e. has a unique s.c.s.

Stage 2 (Determine if A is aperiodic).
We determine the cyclicity of the matrix $\tilde{A}$ defined above. We apply an algorithm of Denardo, [55]. This algorithm uses a covering tree of the graph which is precisely the tree obtained with Tarjan's algorithm.

By definition, matrix A is aperiodic if and only if the cyclicity of $\tilde{A}$ is 1. From now on, we assume it is the case. When it is not the case, the set of eigenelements of matrix A is not projectively bounded. See for example Chapter 6, Example 6.9.5.

Stage 3 (Compute the eigenvalue of A).
We compute λ, the eigenvalue of A by applying an algorithm of Karp [92], see also [78] or [44].

We normalize matrix A, i.e. we set $A := A - \lambda$, i.e. $\forall i, j,\ A_{ij} := A_{ij} - \lambda$.

Stage 4 (Compute all the eigenvectors of A).
By Theorem 3.4.6, we know that it is sufficient to compute matrix $A^+ = A \oplus \cdots \oplus A^k$ to obtain all the extremal eigenvectors. Practically, we are going to compute $A^* = (E \oplus A)^+$ and use $A^+ = AA^*$. We apply an algorithm of Floyd [64] to compute A^* .

The critical columns of A^+ are the extremal eigenvectors. Let us denote by $u^1, \dots u^p$ the different extremal eigenvectors. The complete set of eigenvectors is (see [43] [44]).

$$\{\alpha_1 u^1 \oplus \alpha_2 u^2 \oplus \cdots \oplus \alpha_p u^p,\ \alpha_i \in \mathbb{R}_{max}\}\,. \tag{5.1}$$

We have obtained the eigenvectors but not the periodic regimes of matrix A. Let us illustrate this on an example. We consider

$$A = \begin{pmatrix} -1 & -1 & e \\ e & -1 & -1 \\ -1 & e & -1 \end{pmatrix} \Rightarrow A^+ = \begin{pmatrix} e & e & e \\ e & e & e \\ e & e & e \end{pmatrix}\,.$$

Matrix A is scs1-cyc3, see §3.5.2, Figure 3.6. In order to obtain the extremal periodic regimes, it is necessary to compute $(A^3)^+$ or more generally $(A^l)^+$ where l is the cyclicity (Definition 3.4.5 and Theorem 3.4.7) of A.

Stage 5 (Determine the spectral type of A).
We need to determine the critical graph of A (Definition 3.4.5). Let $u = (u_1, \dots u_k)'$ be one of the eigenvectors of A as computed above. We consider

$$A := P^{-1}AP, \quad \text{where } P = \begin{pmatrix} u_1 & \varepsilon & \varepsilon \\ \varepsilon & \ddots & \varepsilon \\ \varepsilon & \varepsilon & u_k \end{pmatrix}\,. \tag{5.2}$$

All critical terms of A are now equal to e and all non-critical terms are less or equal to e, see Lemma 3.4.15. With this trick, we can now consider the boolean "critical" matrix associated with A and defined as $\tilde{A}_{ij} = \varepsilon$ if $A_{ij} < e$, $\tilde{A}_{ij} = e$ if $A_{ij} = e$. We apply Tarjan's algorithm, see Stage 1. Let p be the number of m.s.c.s of $\tilde{A}$. We apply Denardo's algorithm to each s.c.s., see Stage 2. Let $l_1, \dots, l_p$ be the cyclicity of the different s.c.s. The spectral type of matrix A is scs p-cyc l where $l = \mathrm{lcm}(l_1, \dots, l_p)$.

Stage 6 (Compute all the eigenelements of A).
We compute matrix A^l. We compute matrix $(A^l)^+$, see Stage 4. The critical columns of $(A^l)^+$ are the extremal eigenelements. We denote by $v^1, \dots, v^q$, the different extremal eigenelements. From Theorem 3.4.6, we deduce that $q = \sum_{i=1}^{p} l_i$.
Given an extremal element v^j, the period of the periodic regime $(v^j, Av^j, A^2v^j, \dots)$ is the cyclicity of the corresponding s.c.s. of $\tilde{A}$ as defined in Stage 5.
The complete set of eigenelements of A is

$$\{\alpha_1 v^1 \oplus \alpha_2 v^2 \oplus \cdots \oplus \alpha_q v^q,\ \alpha_i \in \mathbb{R}_{max}\}\,. \tag{5.3}$$

Remark 5.1.1. In the first two stages, we check some structural properties of the graph associated with A. The algorithms considered were proposed more than twenty years ago for general graphs. At that time, the motivation was the study of non-negative matrices (in the usual algebra). The algorithm of Stage 3 was originally proposed in the general framework of

valued graphs.
The problem of computing longest paths in a graph is very classical, see Gondran and Minoux for a complete review, [78] Chapters 2 and 3. The algorithm used in Stage 4 was proposed for this problem. When translated using (max,+) notations, it is exactly equivalent to the computation of A^*.

5.2 Complexity

Here are the PRAM complexities of the different algorithms used in §5.1.

1. $O(k^2)$. See Tarjan [132].
2. $O(k^2)$. See Denardo [55].
3. $O(k^3)$. See Karp [92].
4. $O(k^3)$. See Floyd [64].
5. $O(k^2)$. The computation of Equation (5.2) involves exactly two additions for each term of the matrix. For the remaining, see Stages 1 and 2 above.
6. $O(k^3\sqrt{k\log k})$. Classically, the complexity of computing matrix A^l is $O(k^3 \log l)$. Let us denote by $l(k)$ the maximal possible value of the cyclicity l of a matrix of dimension $k \times k$. The asymptotic of $l(k)$ when $k \to \infty$ is known, see Miller [110] for a survey paper on the subject. We have $\log l(k) \sim \sqrt{k \log k}$. We conclude that the complexity of the computation of A^l is $O(k^3\sqrt{k\log k})$. For the computation of $(A^l)^+$, see Stage 4.

Remark 5.2.1. If one wants to compute the set of eigenelements (5.3) but not the set of eigenvectors (5.1), then it is possible to replace Floyd's algorithm in Stage 4 by an algorithm of Moore [111]. With this algorithm, only one column of matrix A^+ is computed. Moore's algorithm has a complexity $O(k^2)$ to be compared with the complexity $O(k^3)$ of Floyd's algorithm.

5.3 Alternative Algorithm

In [40], Chou and Duffin obtain an eigenvector of matrix A as the solution of two linear programs. In [29], Braker and Olsder propose yet another algorithm working only in the scs1-cyc$l$ case. We are going to study this algorithm more closely.

We consider an irreducible matrix A of size $k \times k$. We suppose that A is scs1-cyc$l$ ($l \leqslant k$). For simplicity of notations, we suppose that A is normalized and that the critical circuit of A is $(1, 2, \ldots, l, 1)$ (we can obtain this just by considering a permutation of the coordinates, see Lemma 3.4.13). The goal is to find its unique eigenvector u. An easy way to do so, is to compute $A^+ = \bigoplus_{n \geqslant 1} A^n$, see Stage 4. It might be seen as the (max,+) translation of the power algorithm of the traditional linear algebra, see [76]. Oddly enough, it is also possible in some cases to compute the eigenvector by applying directly the power algorithm without translating

it in the (max,+) algebra. It is this phenomenon, discovered by Braker and Olsder, which we are going to illustrate using the graphical representation of Chapter 3.

Here is the algorithm.

1. Take an initial vector $x_0 \neq (\varepsilon, \varepsilon, \ldots, \varepsilon)'$.
2. Compute the smallest integer m such that $A^{m+l}x_0 = A^m x_0$, for some $l \in \mathbb{N}$.
3. Consider $v = (A^m x_0 + A^{m+1}x_0 + \cdots + A^{m+l-1}x_0)/l$ (division in the conventional algebra).

In some cases, the vector v is the eigenvector of matrix A. Here are two examples. Figure 5.1 corresponds to matrices :

$$A = \begin{pmatrix} . & e & . \\ . & . & e \\ e & . & . \end{pmatrix}, \; B \begin{pmatrix} . & e & . \\ e & . & . \\ e & . & . \end{pmatrix}, \; (.) = -1 \, .$$

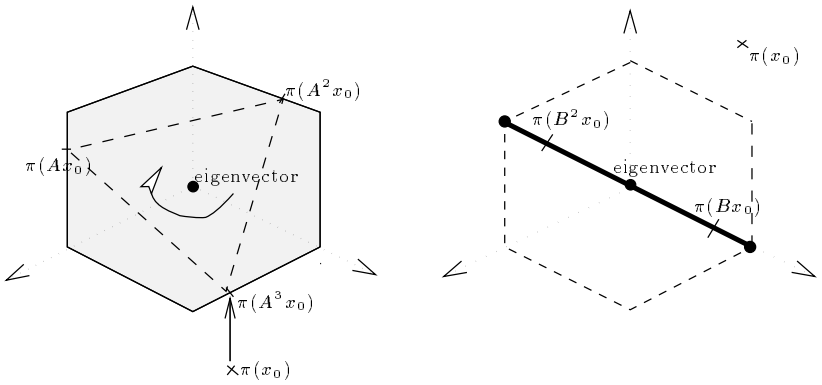


Figure 5.1: scs1-cyc3 and scs1-cyc2, matrices A and B. The algorithm of Braker and Olsder is successful.

Matrix A is scs1-cyc3, its unique eigenvector is $e = (e, e, e)'$. For any periodic regime of period 3, $\{u_1, u_2, u_3\}$, we have $\pi(e) = \pi((u_1 + u_2 + u_3)/3)$.

Matrix B is scs1-cyc2, its unique eigenvector is $e = (e, e, e)'$. We see on Figure 5.1 that the algorithm still works.

Proposition 5.3.1. *The first l components, $v_1, \ldots, v_l$, of the vector v of the algorithm are the correct eigenvector components. However, vector v is not always the eigenvector of A.*

To determine whether v is the eigenvector of A, one has to check if $Av = v$. When v is not the eigenvector of A, Braker and Olsder propose another algorithm which they call the extended algorithm.

1. Define the vector $\hat{v}$ in the following way :

$$\begin{aligned} \hat{v}_i &= v_i \;\; if \;\; (Av)_i = v_i \, , \\ \hat{v}_i &= \varepsilon \;\; if \;\; (Av)_i \neq v_i \, . \end{aligned}$$

2. Let m be the smallest integer such that $A^{m+1}\hat{v} = A^m\hat{v}$.

The integer m is finite and $A^m\hat{v}$ is the unique eigenvector of A.

There exist a technical criterion to determine if the algorithm of Braker and Olsder is going to work or if the extended algorithm is needed, see [29]. But the graphical representation in the projective space gives a very simple illustration of the phenomenon.

Here is now an example where the use of the extended algorithm is necessary.

$$C = \begin{pmatrix} . & e & . \\ e & . & . \\ . & . & -2 \end{pmatrix}, \ (.) = -1 \ .$$

We have represented this example on Figure 5.2.

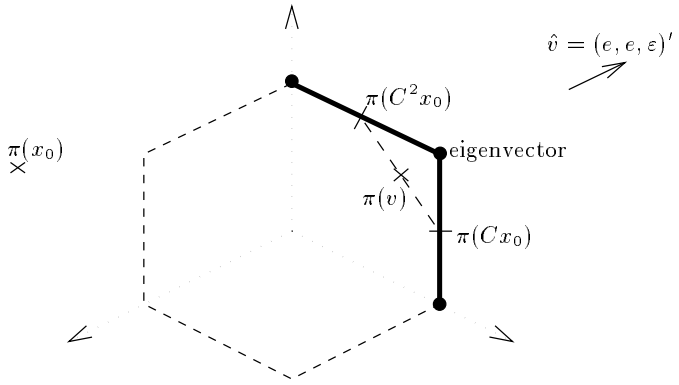


Figure 5.2: scs1-cyc2, matrix C. The algorithm of Braker and Olsder fails.

Matrix C is scs1-cyc2 and its unique eigenvector is $(e, e, -1)'$. By the algorithm of Braker and Olsder, we obtain $v = (e, e, e)'$. With the extended algorithm, we get $C \otimes (e, e, \varepsilon)' = (e, e, -1)'$.

By Proposition 5.3.1, we have that for all matrices of size $k \times k$ which are scs1-cyck, the algorithm of Braker and Olsder provides the eigenvector. For matrices which are scs1-cyc$l$ $(1 < l < k)$, the extended algorithm will, in general, be needed.

In terms of complexity, the algorithm of Braker and Olsder is not optimal. Indeed, one has to compute the stationary version A^m of the matrix and we know that a transient regime can be arbitrarily long, see Section 3.5.5. If we suppose that matrix A is stationary, the complexity is of order $O(k^4)$.

Let us mention that we use Proposition 5.3.1 in a completely different framework in Chapter 4, Proposition 4.4.4.

Partie II

Systèmes Linéaires Stochastiques

Chapitre 6

Products of Random Matrices in the (max,+) Algebra

Produits de Matrices Aléatoires dans l'Algèbre (max,+)

On étudie les systèmes (max,+) linéaires stochastiques. Le problème considéré est celui de la stabilité de tels systèmes. Les résultats proposés permettent de compléter les résultats obtenus par Baccelli [4] et Cohen [46].

Ce chapitre est une adaptation de l'article [105]. Cet article a été accepté pour publication dans *Advances in Applied Probability*. Nous remercions François Baccelli, Serguei Foss, Stéphane Gaubert et un rapporteur anonyme pour de nombreuses remarques et suggestions ayant grandement contribué à l'amélioration de cet article.

We consider a class of closed systems with synchronization, blocking and/or fork-join properties. The main subclass of interest consists in Stochastic Event Graphs. They include cyclic Jackson Networks, many manufacturing models, models with general blocking (such as kanban) and some interacting particle systems. Other models which fit into our framework include task graphs and task graphs with random precedences. For more details on modelling aspects, see Chapter 1. The common feature of these systems is that they can be represented by a linear recursive equation in the (max,+) algebra. We are interested in stationary regimes for quantities such as queue length, waiting times or idle times.

6.1 Introduction

Let us consider the following recursive equation:

$$\left\{ \begin{array}{rcl} x_i(n+1) & = & \max_j(A_{ij}(n) + x_j(n)) \\ x_i(0) & = & (x_0)_i \end{array} \right. \tag{6.1}$$

The sequences $\{A_{ij}(n), i,j = 1,\ldots,k\}$ are given (exogenous data). The process we want to study is the sequence of vectors $\{x(n) = (x_1(n),\ldots,x_k(n))'\}$. The vector x_0 is the initial condition.

It is very fruitful to use a matrix-vector notation for Equation (6.1). We define the following "(max,+)" notations:

$$\varepsilon = -\infty,\ \forall x,y \in \mathbb{R} \cup \{\varepsilon\},\ \ x \oplus y = \max(x,y),\ x \otimes y = x + y\ .$$

We define also the $k \times k$ matrix $A(n) = \{A_{ij}(n), i,j = 1,\ldots,k\}$ and the column vector $x(n) = (x_1(n),\ldots,x_k(n))'$. With these notations, the basic Equation (6.1) takes a very simple and convenient form. In fact, it can be rewritten as:

$$x(n+1) = A(n) \otimes x(n)\ . \tag{6.2}$$

The matrix-vector product is defined in a natural way just by replacing $+$ and $\times$ by $\oplus$ and $\otimes$, i.e. $(A \otimes x)_i = \max_j(A_{ij} + x_j) = \bigoplus_j A_{ij} \otimes x_j$.

We are interested in stochastic versions of Equation (6.2), where $\{A_{ij}(n)\}$ is a sequence of random matrices. As a consequence, here is an equivalent way of introducing our subject: it is a counterpart of the classical theory of products of random matrices (see Furstenberg & Kesten [66] or Bougerol & Lacroix [27]) but in another algebraic structure, the (max,+) algebra.

For systems described by Equation (6.2), we will consider two kinds of asymptotic results.

- First order limits, on ratios:

$$\lim_n \frac{\|x(n)\|_\infty}{n}\ ,\ \ \lim_n \frac{x_i(n)}{n}\ .$$

- Second order limits, on differences:

$$\lim_n\ x_i(n+1) - x_i(n),\ \forall i,\ \ \lim_n\ x_j(n) - x_i(n),\ \forall i \neq j\ .$$

A first order limit is a cycle time or equivalently the inverse of a throughput. Second order limits are related to waiting and idle times, workload, queue length and frequency of occupation. More insights on the relations between these limits and quantities of interest for the system will be provided in Section 6.2, see also Chapter 1. Our goal is to find stationary regimes for second order limits. Multiple stationary regimes will mean multiple possible regimes for waiting times or queue lengths, depending on the initial condition.

Among the systems modeled by Equation (6.2), we can distinguish two classes: the open (or non-autonomous) systems and the closed (or autonomous) ones. Open systems have been exhaustively treated by Baccelli [4] [8] (for both first and second order limits). Problems of existence and uniqueness of first order limits for closed systems have been solved by Cohen [46] (see also [4]). These results are recalled in §6.7.1. This paper deals with the open question of existence and uniqueness of second order limits for closed systems. These problems were considered in several earlier papers (Resing, de Vries, Hooghiemstra, Keane & Olsder [122] and [115], Baccelli [4]) but only sufficient conditions of uniqueness were known. The approach we use is new and exploits completely the common hidden algebraic structure of the different models we consider. It enables us to obtain necessary and sufficient conditions for stability (in some cases) together with simple proofs.

The conditions we give are based on the structure of deterministic matrices chosen in the support of the random matrix $A(0)$. The main result states that the system has a unique stationary regime if the support of $A(0)$ contains a finite number of matrices, $\{A_1, \ldots, A_p\}$, such that the product $A_p \otimes \cdots \otimes A_1$ has a unique periodic regime. The proof makes use of Borovkov's theory of renovating events, see Borovkov & Foss [25] [26]. This theory appears to be much more tractable than classical Harris regeneration due to the specific form of our recursive equations. More details on this last remark are given at the end of Section 6.6.

In order to motivate the practical interest of this work, we present a specific model, a closed cyclic Jackson Network. We are going to use this example throughout the paper to illustrate the theoretical results.

The paper is organized as follows. We introduce two models in Section 6.2, cyclic Jackson Networks and task graphs with random precedences. Sections 6.3, 6.4, 6.5 and 6.6 are presenting the tools that we are using in the paper. They can be skipped by people knowing the subject. Section 6.3 is devoted to the (max,+) algebra, Section 6.4 to the spectral theory in this algebra, Section 6.5 to semigroup of matrices and Section 6.6 to Borovkov's theory of renovating events. Section 6.7 presents the main results. In 6.7.1, we recall some results from [4] and [46]. In 6.7.2, we state some preliminary results. In 6.7.3 and 6.7.4, we give sufficient conditions for the stability of discrete and general models respectively. In Section 6.8, we establish the converses of the results of the previous section. In Section 6.9, we weaken the assumptions under which some of our results apply and we discuss the boundedness of the stationary regime. Finally, for convenience, some of the proofs are given in Appendix.

6.2 Two Motivating Models

6.2.1 Task graphs

We consider a parallel program executed on several identical processors. We model it by its precedence graph τ. If we consider a system of k processors, the graph τ has a set of nodes which is $k \times \mathbb{N}$. The node (i,n) represents the n-th task to be executed at processor i. The arcs between nodes represent the synchronization constraints. There is an arc between the node (i,n) and the node (j,m) (notation : $(i,n) \to (j,m)$) if the n-th task at processor i has to be completed in order for the m-th task at processor j to be enabled. The execution of a task begins as soon as all the tasks of its incoming arcs are completed. Each task has a duration which may depend on the processor.

Let us consider a task graph with synchronizations only between consecutive levels n (i.e. nodes $(1,n),\ldots,(k,n)$) and $n+1$. We assume that the synchronizations depend on n. We denote by $\mathcal{L}(i,n)$ the set of nodes j such that $(i,n) \to (j,n+1)$. We suppose that $\forall i$, there exists a probability law P^i on the subsets of $(1,\ldots,k)$ such that $\mathcal{L}(i,n) = (j_1,\ldots,j_p)$ with probability $P^i\{(j_1,\ldots,j_p)\}$. We denote by $x_i(n)$ the date of completion of task n at processor i, and by $A_{ji}(n)$ the duration of the synchronization constraint between nodes (i,n) and $(j,n+1)$ (it may include a transmission time as well as the execution time at processor j). We adopt the convention that $A_{ji}(n) = -\infty$ if $j \notin \mathcal{L}(i,n)$. It is easy to check that such a model, we could call it a task graph with random precedences, verifies Equation (6.2).
A Queuing Network model studied by Baccelli & Liu [14] corresponds to this model. It is a Kelly type Network (i.e. routes are associated with customers) with a locally FIFO priority rule. The task resource models to be studied in Chapter 9 also have this kind of structure.

6.2.2 Cyclic Jackson network

We consider a closed Jackson Network. The study of such closed networks can be traced back to Gordon and Newell, [79]. In their original model, there is a given number of indistinguishable customers. The routing of the customers leaving a given queue is provided by a sequence of i.i.d. Bernouilli random variables. All the service times are exponential. They prove the existence of an explicit product form for the unique stationary distribution.

A natural generalization of the basic model is to consider i.i.d. service times with general distributions, i.e. to replace $./M/1/\infty$ servers by $./GI/1/\infty$ servers. Finding the minimal assumptions leading to a unique stationary regime for this generalized closed Jackson Network is still an open problem.

We consider a restriction of the previous model. There are k queues and all customers have the same cyclic route $(1,2,\ldots,k,1)$, see Figure 6.1. We will denote this model by CJN for Cyclic Jackson Network, following the terminology of [94].
In the following, the numbering of queues has to be understood modulo $[k]$, for example queue $(k+2)$ is queue 2. We denote by $\{\sigma_j(n), n \in \mathbb{N}\}$, the sequence of service times at queue j. This sequence is i.i.d. We suppose also that the service times at the different queues are independent. Instead of describing the system by the workload or the queue length process, as is usually done, we propose to study this model by introducing the following variables. With each queue j, we

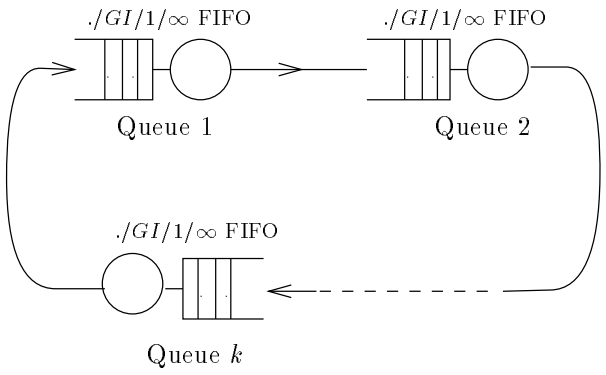


Figure 6.1: A Cyclic Jackson Network consisting of k queues.

associate a dater $\{x_j(n), n \in \mathbb{N}\}$. The variable $x_j(n)$ represents the date of completion of the $n-$th service at queue j. All variables of interest for the network can be derived from these daters and from the sequences of service times. More precisely, we have:

- Asymptotic throughput at queue j:

$$\gamma_j = \lim_{n \to +\infty} \frac{n}{x_j(n)} .$$

- Idle time of queue j before the arrival of the n-th customer to visit queue j.

$$I_j(n) = x_j(n) - \sigma_j(n) - x_j(n-1) .$$

- Workload at queue j at the instant of the arrival of the n-th customer to visit queue j. This customer comes from queue $j-1$. We suppose that it was the n'-th customer to visit queue $j-1$.

$$W_j(n) = x_j(n) - \sigma_j(n) - x_{j-1}(n') .$$

The variables (γ_i) which are obtained as ratios of daters will be called first order variables. The ones (I_j, W_j) which are obtained as differences of daters will be called second order variables. We want to derive conditions under which there is a unique stationary regime for both first and second order variables. In such a case, we say that our model is *stable*.

Suppose for the moment that there are exactly k customers. We suppose also that there is initially one customer in each queue. These assumptions together with the FIFO service discipline at each queue yields the following property. The n-th customer to visit queue j will be, at the next step of its route, the $(n+1)$-th customer to visit queue $j+1$. As a consequence we are

able to write a recursive evolution equation for the daters.

$$\left\{\begin{array}{lcl} x_1(n+1) & = & \max(x_1(n), x_k(n)) + \sigma_1(n) \\ x_2(n+1) & = & \max(x_2(n), x_1(n)) + \sigma_2(n) \\ & \cdots & \\ x_j(n+1) & = & \max(x_j(n), x_{j-1}(n)) + \sigma_j(n) \\ & \cdots & \end{array}\right. .$$

Using the (max,+) notation this can be rewritten as:

$$x(n+1) = A(n) \otimes x(n), \text{ where } A(n) = \begin{pmatrix} \sigma_1(n) & \varepsilon & \cdots & \varepsilon & \sigma_1(n) \\ \sigma_2(n) & \sigma_2(n) & \ddots & & \varepsilon \\ \varepsilon & \ddots & \ddots & \varepsilon & \vdots \\ \vdots & \ddots & \ddots & \ddots & \varepsilon \\ \varepsilon & \cdots & \varepsilon & \sigma_k(n) & \sigma_k(n) \end{pmatrix} . \qquad (6.3)$$

The initial condition is $x(0) \geqslant 0$, where $x_i(0)$ is the remaining service time of the customer being served at queue i at time 0.

When the service times are deterministic, it is possible to obtain many asymptotic behaviors, depending on the initial condition $x(0)$. In fact, initial delays between customers might never vanish. Therefore, it is possible to have several stable regimes for second order quantities $(I_j, W_j, \ldots)$ including periodic ones. For stochastic systems, when the service times are random variables, it is still possible to have several stationary regimes if the system is not "stochastic enough". As an application of the results presented in this paper, we obtain the necessary and sufficient conditions for the existence of a unique stationary regime for this CJN. This basic CJN will be used as an illustration of the results throughout the paper (Examples 6.4.8, 6.7.9, 6.7.16 and 6.8.7).

When there are less than k customers in the network, the system can be represented in the same way as previously. The only difference is that the structure of matrices $\{A(n)\}$ is more complicated. When there are more than k customers, the trick consists in splitting queues. Each queue which has originally $(p > 1)$ customers in its buffer is transformed into p queues with one customer per buffer. This is done by creating $p-1$ fictive queues with service times identically equal to zero. By doing this, one gets back to the previous case. The main difference is that we have represented our model by a (max,+) linear system of dimension greater than the original number of queues. For more details on these transformations, see Chapter 1, §1.3.

Many generalizations of this basic CJN can be made within the class of systems admitting a (max,+) linear representation. For example, we can consider queues with stationary and ergodic sequences of service times (the $./G/1/\infty$ case). The modeling is exactly the same. The only difference is that the sequence of random matrices $\{A(n)\}$ is stationary-ergodic instead of being i.i.d. Our results apply to the stationary-ergodic framework. We can also consider finite buffers (the $./G/1/L$ case), see §1.3.3 for details. In the case of a CJN with i.i.d. general service times (./GI/1 servers), there is an alternative method for studying the network. We consider the Markov chain formed by queue lengths and remaining service times, and we apply Harris

regeneration techniques.
On the one hand, it is possible to obtain natural sufficient conditions of stability. Consider the configuration where all the customers are blocked at the same queue. If this configuration is of positive probability, it can be used as a regeneration point for the Markov chain. To obtain a positive probability, it is enough to have one of the service times with an unbounded support. This kind of ideas was first introduced for closed acyclic Jackson Networks by Borovkov, [23], [24]. For closed Cyclic Jackson Networks, this technique, with some refinements, is used by Bambos [16] and Kaspi & Mandelbaum [94].
On the other hand, obtaining necessary and sufficient conditions of stability is a difficult task. As far as we know, it has been done only in the case of the basic Cyclic Jackson Network with two queues, see [94]. For the basic Cyclic Jackson Networks with N queues, the best sufficient conditions that we have found in the literature are given in [95], see Example 6.8.7. In this paper, the authors show that, in some cases, there is stability of a CJN even if the configurations where all the customers are blocked at the same queue never happen.

Our approach allows us to derive the necessary and sufficient conditions of stability, for all the CJN mentioned above. The drawback is that it requires a preliminary modeling stage, the translation of the "real system" into its (max,+) linear representation.

6.3 (max,+) Algebra

Definition 6.3.1 ((max,+) algebra). *We consider the semiring* $(\mathbb{R}\cup\{-\infty\},\oplus,\otimes)$. *The law* $\oplus$ *is "max" and* $\otimes$ *is the usual addition. We set* $\varepsilon=-\infty$ *and* $e=0$. *The element* ε *is neutral for the operation* $\oplus$ *and absorbing for* $\otimes$. *The element* e *is neutral for* $\otimes$. *The law* $\oplus$ *is idempotent, i.e.* $a\oplus a=a$. $(\mathbb{R}\cup\{\varepsilon\},\oplus,\otimes)$ *is an idempotent semiring, called a dioid. We shall denote it by* $\mathbb{R}_{max}$.

In the rest of the paper, the notations "$+,\times$" will stand for the usual addition and multiplication. Nevertheless, we will write ab for $a\otimes b$ whenever there is no possible confusion. For example, for $a\in\mathbb{R}$, $a^d=a^{\otimes d}=d\times a$.

We define the product spaces $\mathbb{R}_{max}^k$, $\mathbb{R}_{max}^{k\times k}$. We define the product of a vector by a scalar: $a\in\mathbb{R}_{max}, u\in\mathbb{R}_{max}^k$, $(a\otimes u)_i=a\otimes u_i$.

Matrix product is defined in a natural way, replacing $+$ and $\times$ by $\oplus$ and $\otimes$ respectively. Let $A,B\in\mathbb{R}_{max}^{k\times k}$,

$$(A\otimes B)_{ij}=\max_l(A_{il}+B_{lj})=\bigoplus_l A_{il}\otimes B_{lj}\ .$$

Matrix-vector product is defined in a similar way.

Let us recall some definitions adapted to the $\mathbb{R}_{max}$ algebra.

Definition 6.3.2. *The (communication) graph of a square matrix* A *is a directed graph having a number of nodes equal to the size of* A. *This graph contains an arc from* i *to* j *iff* $A_{ji}\neq\varepsilon$. *The valuation of this arc is* A_{ji}.

Definition 6.3.3. *A square matrix A is irreducible if:* $\forall i,j \quad \exists n \geqslant 0 \quad | \quad (A^n)_{ij} > \varepsilon$ *(or equivalently if its communication graph is strongly connected).*

Definition 6.3.4. *A square matrix A is aperiodic if:* $\exists N,\ \forall n \geqslant N,\ \forall i,j,\quad (A^n)_{ij} > \varepsilon$.

Definition 6.3.5. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A stochastic matrix $\{A(\omega),\ \omega \in \Omega\}$ has a fixed structure if $P(A_{ij} = \varepsilon) = 1$ or $P(A_{ij} = \varepsilon) = 0$, $\forall i,j$.*

Definition 6.3.6 ($\mathbb{PR}^k$). *The projective space $\mathbb{PR}^k$ is defined as the quotient of $\mathbb{R}^k$ by the parallelism relation:*

$$u, v \in \mathbb{R}^k \quad u \simeq v \iff \exists a \in \mathbb{R} \text{ such that } u = a \otimes v \ .$$

Let π be the canonical projection of $\mathbb{R}^k$ into $\mathbb{PR}^k$.

For example $(e, -1)'$ and $(2,1)' = (e+2, -1+2)'$ are in the same parallelism class, i.e. are two representatives of the same element of $\mathbb{PR}^k$. We define in the same way $\mathbb{PR}^k_{max}$, $\mathbb{PR}^{k\times k}_{max}$ and $\mathbb{PR}^{k\times k}$. We use the same notation π for the different canonical projections. We define a norm and a distance on $\mathbb{PR}^k$ which we are going to call the projective norm and distance.

Definition 6.3.7. *Let $x \in \mathbb{PR}^k$ and $u \in \mathbb{R}^k$ be a representative of x, i.e. $\pi(u) = x$. We define:*

$$|x|_{\mathcal{P}} = \max_i u_i - \min_i u_i \ .$$

Let $x, y \in \mathbb{PR}^k$ and $u, v \in \mathbb{R}^k$ be two representatives of x and y respectively. We define:

$$d(x,y) = d(u,v) = |x - y|_{\mathcal{P}} = \bigoplus_i (u_i - v_i) \otimes \bigoplus_i (v_i - u_i) \ .$$

We write either $d(x,y)$ or $d(u,v)$ with some abuse of notation.

The space $(\mathbb{PR}^k, |.|_{\mathcal{P}})$ is an Euclidean space. In particular, it is complete. Indeed, it is easy to check that $|x|_{\mathcal{P}}$ does not depend on the representative u, and is a norm on $\mathbb{PR}^k$, viewed as a vectorial space on $\mathbb{R}$. This norm corresponds to the $\mathcal{L}_\infty$ norm[1] on the projective space $\mathbb{PR}^k$. We have the following very important property.

Proposition 6.3.8. *Let $A \in \mathbb{R}^{k\times k}_{max}$ be an irreducible matrix. Let u, v be two vectors of $\mathbb{R}^k$. We have:*

$$d(Au, Av) \leqslant d(u,v) \ .$$

[1] It is worth mentioning that $d(.,.)$ is the $\mathbb{R}_{max}$ analogue of a distance used in classical Perron-Frobenius theory, which is called the Hilbert's projective metric and is defined by "$\delta(u,v) = \ln\left(\inf\{\mu/\lambda \mid \lambda u \leqslant v \leqslant \mu u\}\right)$".

Proof. By definition, we have:

$$d(Au,Av) = \bigoplus_i ((Au)_i - (Av)_i) \otimes \bigoplus_i ((Av)_i - (Au)_i) .$$

We define $j(i)$ such that $(Au)_i = \bigoplus_j A_{ij} \otimes u_j = A_{ij(i)} \otimes u_{j(i)}$. Note that $j(i)$ depends on A and u. We have:

$$\begin{aligned}
\bigoplus_i ((Au)_i - (Av)_i) &= \bigoplus_i \left((\bigoplus_j A_{ij} \otimes u_j) - (\bigoplus_j A_{ij} \otimes v_j) \right) \\
&= \bigoplus_i \left((A_{ij(i)} \otimes u_{j(i)}) - (\bigoplus_j A_{ij} \otimes v_j) \right) \\
&\leqslant \bigoplus_i (A_{ij(i)} \otimes u_{j(i)} - A_{ij(i)} \otimes v_{j(i)}) \\
&= \bigoplus_i u_{j(i)} - v_{j(i)} \leqslant \bigoplus_i u_i - v_i
\end{aligned}$$

We obtain $d(Au,Av) \leqslant \bigoplus_i (u_i - v_i) \otimes \bigoplus_i (v_i - u_i)$ i.e. $d(Au,Av) \leqslant d(u,v)$. □

There is no simple criterion to get a strict inequality. This monotonicity has to be interpreted as a synchronization property.

Definition 6.3.9. *We consider $A \in \mathbb{R}_{max}^{k \times k}$. We set*

$$\mathtt{D}(A) = \sup_{u,v \in \mathbb{R}^k} d(Au,Av) .$$

We call $\mathtt{D}(A)$ the projective diameter of A.

In order for the previous definition to be non-ambiguous, it is necessary that $\forall u \in \mathbb{R}^k, Au \in \mathbb{R}^k$. It implies that $\forall i, \exists j$ s.t. $A_{ij} > \varepsilon$. It is easy to prove that $\mathtt{D}(A)$ is finite if and only if $\forall i,j,\ A_{ij} > \varepsilon$. A matrix A can be considered as a "linear" (in the (max,+) sense) operator from $\mathbb{PR}^k$ into $\mathbb{PR}^k$. It is a bounded operator if the (decreasing) sequence $\mathtt{D}(A^n)$ has a finite limit, i.e. if A is aperiodic (Def. 6.3.4).

6.4 Deterministic Spectral Theory

We recall some results of the deterministic spectral theory in the $\mathbb{R}_{max}$ algebra. For references, see Section 3.4.

We want to find non trivial solutions to the *eigenvalue problem* :

$$A \otimes x = \lambda \otimes x ,$$

where $A \in \mathbb{R}^{k \times k}$ is an irreducible matrix, x is a column vector (the "eigenvector") and λ is a real constant (the "eigenvalue"). We define also periodic regimes for the eigenvalue problem. A periodic regime of period d is a set of vectors $\{x_1, \ldots, x_d\}$ of $\mathbb{R}^k$ verifying $Ax_i = \lambda x_{i+1}$, $i = 1, \ldots, d-1$ and $Ax_d = \lambda x_1$.

Definition 6.4.1. *For each path $\zeta = \{t_1, t_2, \cdots, t_j, t_{j+1} = t_1\}$, we define its average weight by:*

$$p(\zeta) = \frac{a_{t_1 t_j} \otimes \cdots \otimes a_{t_3 t_2} \otimes a_{t_2 t_1}}{j} ,$$

(here the division is the conventional one).

Theorem 6.4.2. *There is a unique (non ε) eigenvalue, λ. It satisfies the relation*

$$\lambda = \max_{\zeta} p(\zeta) ,$$

where ζ covers all the circuits of (the communication graph of) A. We call also λ the Lyapunov exponent or the cycle time of A.

There might be several eigenvectors. A linear combination (in $\mathbb{R}_{max}$) of eigenvectors is an eigenvector. An eigenvector has all its coordinates different from ε (due to the irreducibility assumption).

Definition 6.4.3. *For a matrix A, we define:*

Critical circuit *A circuit ζ of A is said to be critical if its average weight is maximal, i.e. if $p(\zeta) = \lambda$.*

Critical graph *It consists of the nodes and arcs of A belonging to the critical circuit(s).*

For a general graph, we define :

Cyclicity *The cyclicity of a strongly connected graph is the greatest common divisor of the lengths of all the circuits. The cyclicity of a connected graph is the least common multiple of the cyclicities of its maximal strongly connected subgraphs (s.c.s.).*

We normalize a matrix by subtracting (in the conventional algebra) the eigenvalue to all the coordinates. The eigenvalue of a normalized matrix is e. For a normalized matrix A of size k, we define:

$$A^+ = A \oplus A^2 \oplus \cdots \oplus A^k .$$

We check that $A^+ \oplus A^{k+1} = A^+$.

Theorem 6.4.4. *Let A be a normalized matrix.*

a. Critical columns $A^+_{\cdot i}$, i belonging to the critical graph, are eigenvectors.

b. For i, j belonging to the critical graph, $\pi(A^+_{\cdot i})$ and $\pi(A^+_{\cdot j})$ are different iff they belong to two different s.c.s. of the critical graph.

c. Every eigenvector of A writes as a linear combination (in $\mathbb{R}_{max}$) of critical columns $A^+_{\cdot i}$.

A corollary of this result will be of particular use for us:
An irreducible matrix has a unique eigenvector (up to a multiplicative ($\otimes$) constant) if and only if its critical graph has a unique s.c.s.

In $\mathbb{R}_{max}$, every irreducible matrix is cyclic in the sense of the following theorem.

Theorem 6.4.5. *For an irreducible matrix A of size k and whose eigenvalue is λ, there exists integers d and M such that:*

$$\forall m \geqslant M, \quad A^{m+d} = \lambda^{\otimes d} \otimes A^m \,, \tag{6.4}$$

furthermore the smallest d verifying the property is equal to the cyclicity of the critical graph of A. We call it the **cyclicity** *of A.*

A cyclicity greater than one will provide periodic regimes of period greater than one for the eigenvalue problem.

Proposition 6.4.6. *An irreducible matrix has a unique eigenvector and no periodic regimes of period greater than one for the eigenvalue problem, if and only if its critical graph has a unique s.c.s. and its cyclicity is one. Such a matrix will be called a* **scs1-cyc1** *matrix.*

The proof follows from Theorems 6.4.4 and 6.4.5.

Definition 6.4.7 (rank). *By analogy with classical linear algebra, we define the "rank" of a matrix A as the number of additively independent columns (resp. lines) of A. More precisely, let $A_{.i}$ denote the i-th column of A. Matrix A is of rank r if there exists $\mathcal{J} \subset \{1, \ldots, k\}$ such that $|\mathcal{J}| = r$ and $\forall i \neq j \in \mathcal{J}$, $\pi(A_{.i}) \neq \pi(A_{.j})$ and $\forall i \notin \mathcal{J}$, $\exists \alpha_j, j \in \mathcal{J}$, such that*

$$\pi(A_{.i}) = \pi[\bigoplus_{j \in \mathcal{J}} \alpha_j \otimes A_{.j}] \,.$$

Let A be a **rank 1** matrix. Then A is a **scs1-cyc1** matrix and verifies $A^2 = \lambda \otimes A$ (λ is the eigenvalue of A). The other way round, let A be a **scs1-cyc1** matrix and M be defined as in Equation (6.4). One can check that A^M is a matrix of **rank 1**.

Example 6.4.8. [Cyclic Jackson Network 1]

Let us consider a basic Cyclic Jackson Network as presented in Section 6.2.2. We suppose that the service times are deterministic, i.e $\sigma_j(n) \equiv \sigma_j$. We suppose also that the number of customers, k, is equal to the number of queues. Then we can consider the (max,+) matrix associated with the network, see Equation (6.3). The graph associated with this matrix is constituted by the circuit $(1, 2, \ldots, k, 1)$ and the recycling loops $(1, 1)$ to (k, k). Let us define $I = \{i \mid \ \sigma_i = \max_j \sigma_j\}$. There are two possible cases.

- If the cardinal $|I| < k$, then the critical graph of the matrix consists of the nodes $i \in I$ and the arcs $(i, i), i \in I$. It implies that the matrix is scs$|I|$-cyc1.
- If $|I| = k$ then the graph and the critical graph of the matrix coincide. It implies that the matrix is scs1-cyc1.

We conclude that the matrix is scs1-cyc1 if and only if $|I| = 1$ or k.

6.5 Semigroup of Matrices

Definition 6.5.1. *Let us consider $A_1, \ldots, A_p \in \mathbb{R}_{max}^{k\times k}$. We denote by $< A_1, \ldots A_p >$, the semigroup generated by these matrices and by $\pi < A_1, \ldots A_p >$ its projection. We have*

$$\begin{aligned} < A_1, \ldots, A_p > &= \{(A_{u_n} \cdots A_{u_2} A_{u_1}) \,,\ u_1, \ldots, u_n \in \{1, \ldots, p\}, n \in \mathbb{N}\} \,, \\ \pi < A_1, \ldots, A_p > &= \{\pi(A_{u_n} \cdots A_{u_2} A_{u_1}) \,,\ u_1, \ldots, u_n \in \{1, \ldots, p\}, n \in \mathbb{N}\} \,. \end{aligned}$$

Definition 6.5.2. *We say that the semigroup $< A_1, \ldots A_p >$ is primitive if there exists N such that*

$$\forall n \geqslant N,\ u_1, \ldots, u_n \in \{1, \ldots, p\},\ \forall i, j,\ \ (A_{u_n} \cdots A_{u_2} A_{u_1})_{ij} > \varepsilon \,. \tag{6.5}$$

The following result follows from classical arguments.

Proposition 6.5.3. *The semigroup $< A_1, \ldots A_p >$ is primitive if and only if all the matrices of $< A_1, \ldots, A_p >$ are aperiodic (Def. 6.3.4).*

Proof. If one of the matrices, say A, is not aperiodic, then $\forall n, \exists i, j$ such that $(A^n)_{ij} > \varepsilon$. Hence the semigroup is not primitive.
Let us prove the sufficient part of the proposition. This proof was mentioned to me by Stéphane Gaubert (unpublished work). It is enough to prove the result for Boolean matrices, i.e. matrices which coordinates are either e or ε. The only idempotent aperiodic Boolean matrix is the matrix E, $E_{ij} = e, \forall i, j$. In a finite semigroup, there exists N such that all products of length greater than N contain an idempotent, see for example Pin [118]. It implies that long enough products can be written under the form AEB where A and B are aperiodic (as all the matrices of the semigroup are). Matrices of the form AEB have all their coordinates different from ε which concludes the proof. □

We consider the Euclidean space $(\mathbb{PR}^{k\times k}, |.|_{\mathcal{P}})$ as introduced in Definition 6.3.7.

Proposition 6.5.4. *Let $A_1, \ldots, A_p \in \mathbb{Q}_{max}^{k\times k}$. For all compact set K of $(\mathbb{PR}^{k\times k}, |.|_{\mathcal{P}})$, we have $\pi < A_1, \ldots, A_p > \cap K$ is finite. If we assume furthermore that $< A_1, \ldots, A_p >$ is primitive then $\pi < A_1, \ldots, A_p >$ is finite.*

Proof. The second part of the proposition was proved by Gaubert [70]. The first part is obtained by a slight modification of the proof of [70]. □

Proposition 6.5.4 can be extended to matrices such that $\forall i = 1, \ldots, p,\ \exists \lambda_i \in \mathbb{R}$ such that $\lambda_i \otimes A_i \in \mathbb{Q}_{max}^{k\times k}$. But it cannot be extended to matrices in $\mathbb{R}_{max}^{k\times k}$ as shown in Chapter 3, §3.7 or in Example 6.8.2. It is the reason why some of our results apply only for models with matrices in $\mathbb{Q}_{max}^{k\times k}$, see Section §6.8.1.

6.6 Borovkov's Renovating Events Theory

Borovkov's theory deals with the problem of regeneration in so-called "Stochastic Recursive Sequences". For a complete treatment, the reader is referred to Borovkov [22], Borovkov & Foss [25, 26] or Brandt, Franken & Lisek [32]. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let θ be a measurable map from $(\Omega, \mathcal{F})$ into itself such that P is θ-invariant and θ-ergodic. Let $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be two Polish spaces (complete, separable metric spaces) equipped with their respective Borel σ-algebra.

Definition 6.6.1. *We call Stochastic Recursive Sequence (SRS), a sequence $\{x(n)\}$ of E-valued random variables defined by*

$$x(n+1) = f(x(n), a(n)),\ n \geqslant 0,\ x(0) = x_0\ ,$$

where $\{a(n)\}$ is an exogenous sequence of G-valued random variables, stationary with respect to the shift θ. The function f is a measurable function from $E \times G$ into E. The vector $x_0 \in E$ is the initial condition. In order to stress the value of the initial condition, we will sometimes denote the SRS by $\{x(n, x_0)\}$.
We talk of an i.i.d. SRS when the sequence $a(n)$ is i.i.d. (an i.i.d. SRS is a Markov chain and the converse is true!).

Definition 6.6.2. *We consider $\{x(n)\}$, a SRS. We denote by $\mathcal{F}_l$ the σ-algebra $\mathcal{F}_l = \sigma\{a(n),\ n \in \{-\infty, \ldots, l-1\}\}$. The sequence of events $\{\mathcal{A}(n) \in \mathcal{F}_{n+m},\ n \in \mathbb{N}\}$ is said to be a renovating sequence of length m and of associated function $\phi : G^m \to E$ if:*

$$\exists n_0,\ \forall n \geqslant n_0,\ x(n+m) = \phi\left(a(n), a(n+1), \cdots, a(n+m-1)\right) \ \text{ on } \ \mathcal{A}(n)\ .$$

A sequence $\{\mathcal{A}(n), n \in \mathbb{N}\}$ of renovating events of same length and associated function is said to be stationary if $\mathcal{A}(n) = \mathcal{A}(0) \circ \theta^n = \theta^{-n}\mathcal{A}(0)$.

We need the following notions of convergence:

Definition 6.6.3. *We say that there is coupling convergence in finite time (or, merely, coupling) of a sequence $\{X_n\}$ to a stationary sequence $\{Y \circ \theta^n\}$ if*

$$P(X_{n+l} = Y \circ \theta^{n+l},\ \forall l \geqslant 0) \overset{n \to +\infty}{\longrightarrow} 1\ .$$

It is easy to show that this notion of coupling convergence implies total variation convergence ($X_n \to Y$ in total variation if $\sup_{A \in \mathcal{F}} |P(X_n \in A) - P(Y \in A)| \overset{n \to +\infty}{\longrightarrow} 0$).

Definition 6.6.4. *We say that there is strong coupling convergence in finite time (or, merely, strong coupling) of a sequence $\{X_n\}$ to a stationary sequence $\{Y \circ \theta^n\}$ if:*

$$\nu = \min\left\{n \geqslant 0 \mid X_{n+l} \circ \theta^{-(n+l)} = Y,\ \forall l \geqslant 0\right\} \ \text{ is a.s finite}$$

(the sequence $\{X_n \circ \theta^{-n}\}$ corresponds to the famous Loynes scheme).

Remark Strong coupling implies coupling but the converse is not true.

Theorem 6.6.5 (Borovkov's renovating events). *We consider a SRS $\{x(n)\}$ defined by:*

$$x(n+1) = f(x(n), a(n)),\ n \geqslant 0,\ x(0) = x_0\ .$$

If the random process $\{x(n), n \in \mathbb{N}\}$ admits a stationary sequence of renovating events $\{\mathcal{A}(n)\}$ such that $P(\mathcal{A}(0)) > 0$, then there exists a finite random variable Z such that:

$$Z \circ \theta = f(Z, a(0))\ .$$

The sequence $\{Z \circ \theta^n\}$ is a stationary regime for the SRS and $x(n)$ converges with strong coupling in finite time to $Z \circ \theta^n$.

In the previous theorem, we have considered a SRS defined with a unique initial condition, x_0. In the rest of the paper, we will be interested in having results that hold uniformly over the initial conditions. We will then use the following generalization of Borovkov's theorem.

Theorem 6.6.6. *We consider a subset V of E ($V = E$ is in particular possible). We suppose that there exists a stationary sequence of events $\{\mathcal{A}(n)\}$ verifying $P(\mathcal{A}(0)) > 0$ and which is renovating for the SRS $\{x(n, x_0)\}$, $\forall x_0 \in V$. Then, for all (possibly random) initial condition $x(0)$ such that $P(x(0) \in V) = 1$, the sequence $\{x(n)\}$ converges with strong coupling to a unique stationary regime.*

Theorem 6.6.7 (converse of Th. 6.6.5 and 6.6.6). *The conditions of Theorem 6.6.5 are necessary and sufficient for strong coupling convergence. Let V be a compact subset of E. The conditions of Theorem 6.6.6 are necessary and sufficient for strong coupling convergence uniformly over initial conditions in V.*

Next theorem was proved by Anantharam and Konstantopoulos in [2].

Theorem 6.6.8. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. We assume that $(\Omega, \mathcal{F})$ is a Polish space equipped with its Borel σ-algebra. We consider a SRS "$x(n+1) = f(x(n), a(n))$" defined on E. Suppose that, for some $x_0 \in E$, the sequence $\{x(n, x_0)\}$ is tight[2] on E. Then there is a stationary distribution for the SRS.*

The stationary distribution is defined on $\Omega \times E$ with an Ω marginal equal to P. It provides only a *weak stationary regime* (*wsr*) for the SRS, see [2] or [32] for details. All we need to know about *wsr* is that stationary regimes are *wsr*. Hence, the uniqueness of stationary regimes implies the uniqueness of *wsr*.

Remark It is proved in [25], that for an i.i.d. SRS (i.e. Markov chain), the conditions of Th. 6.6.5 are equivalent to the ones ensuring Harris ergodicity. In Harris' framework, the conditions are on the state space. In Borovkov's framework, the conditions are on the exogenous driving sequence. This second approach is better suited for our problem. On the one hand, a direct analysis on the state space appears to be almost inextricable. On the other hand, the renovating events will take a very convenient form because a product of matrices is still a matrix (see Theorems 6.7.8, 6.7.10).

[2] Tightness on E means that for any $\eta > 0$, there is a compact K of E such that $P\{x(n, x_0) \in K\} > 1 - \eta$, for all n.

η-coupling Coupling and strong coupling, introduced above, are related to total variation convergence. We define now the notion of η-coupling. It is related to weak convergence.

Definition 6.6.9 (η-coupling). *We consider a metric space (E,d). We consider two sequences $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ defined on E. We say that there is η-coupling[3] of these two sequences if for each $\eta > 0$, one can find versions of $\{X_n\}$ and $\{Y_n\}$ defined on a common probability space and an a.s. finite random time N such that*

$$n \geqslant N \Longrightarrow d(X_n, Y_n) \leqslant \eta \ .$$

The following proposition is shown in Asmussen [3].

Proposition 6.6.10. *We consider a sequence $\{X_n\}_{n\in\mathbb{N}}$ and a stationary sequence $\{Y\circ\theta^n\}_{n\in\mathbb{N}}$ defined on the metric space E. Let μ be the invariant distribution of Y. If there is η-coupling of the two sequences, then $\{X_n\}$ converges weakly to μ.*

6.7 Presentation of the Results

Let us consider a probability space $(\Omega, \mathcal{F}, P, \theta)$. The probability P is stationary and ergodic with respect to the shift θ. We are interested in systems of the type:

$$\left\{ \begin{array}{rcl} x(n+1) & = & A(n) \otimes x(n), \quad n \in \mathbb{N} \\ x(0) & = & x_0 \end{array} \right.$$

where $x(n)$ and $A(n)$ $(\forall n)$ are finite, respectively $\mathbb{R}_{max}^k$ and $\mathbb{R}_{max}^{k\times k}$-valued, random variables. We are sometimes going to use the notation $x(n, x_0)$ to emphasize the value of the initial condition. We will consider models where the sequences $\{A(n), n \in \mathbb{N}\}$ are respectively **i.i.d** or **stationary and ergodic** (i.e $A(n+1) = A(n) \circ \theta$). We are interested in two kinds of asymptotic limits.

- First order limits, on ratios:

$$\lim_n \frac{\|x(n)\|_\infty}{n} \ , \quad \lim_n \frac{x_i(n)}{n} \ .$$

- Second order limits, on differences:

$$\lim_n \ x_i(n+1) - x_i(n), \ \forall i \ ,$$

$$\lim_n \ x_j(n) - x_i(n), \ \forall i \neq j \ .$$

First order limits and second order limits for open systems have been treated by Baccelli [4]. First order limits for closed systems have been treated by Cohen [46]. We are going to recall these results, before completing the picture by solving the problem of second order results for closed systems.

[3] The classical terminology is ε-coupling. We change it to η-coupling to avoid confusions with the notation $\varepsilon = -\infty$ of the $\mathbb{R}_{max}$ algebra.

6.7.1 Results from Baccelli [4] and Cohen [46]

For $x \in \mathbb{R}^k$ and $A \in \mathbb{R}_{max}^{k\times k}$, we use the notation $\|x\|_\infty = \bigoplus_{i=1}^k x_i$ and $\|A\|_\infty = \bigoplus_{i,j=1}^k A_{ij}$.

First order limits for closed systems

Theorem 6.7.1 (Cohen [46]). *Let $\{A(n)\}$ be a stationary and ergodic sequence of matrices. We suppose that $\forall i,j,\ P(A_{ij}(0) = \varepsilon) = 0$ and $\varepsilon < E(A_{ij}(0)) < +\infty$. There exists a constant $\lambda \in \mathbb{R}$ such that, for all initial condition x_0 and for all $i \in \{1,\dots,k\}$*

$$\lim_n \frac{x_i(n,x_0)}{n} = \lim_n E\left(\frac{x_i(n,x_0)}{n}\right) = \lambda,\ \ P-a.s.$$

The constant λ is called the Lyapunov exponent of the stochastic matrix $A(0)$.

Remark The assumptions of Th. 6.7.1 can be weakened and replaced by :

$$\lim_n P\left(\ A(n)\otimes A(n-1)\otimes\cdots\otimes A(0) \text{ irred.}\ \right) = 1,\ \varepsilon < E(A_{ij}(0) \mid A_{ij}(0) \neq \varepsilon) < +\infty\ .$$

Remark If the matrices are of dimension 1, Theorem 6.7.1 is exactly the Strong Law of Large Numbers.

Remark This definition of a Lyapunov exponent is coherent with the one of Theorem 6.4.2. Indeed, by Theorem 6.4.5, for every irreducible and deterministic matrix A, there exists d and M such that $\forall m \geqslant M,\ A^{m+d} = \lambda^d \otimes A^m$, where λ is the eigenvalue of A. It implies that $\forall x_0 \in \mathbb{R}_{max}^k, \lim_n A^n x_0/n = \lambda$.

Proof. It is straightforward to check that :

$$\forall A,B \in \mathbb{R}_{max}^{k\times k},\ \|A\otimes B\|_\infty \ \ \leqslant\ \ \|A\|_\infty \otimes \|B\|_\infty$$

We define $Z_{l,n} = \|A(n-1)\otimes\cdots\otimes A(l)\|_\infty$, $\forall l < n$. We have for all $l < m < n$,

$$\|A(n-1)\cdots A(l)\|_\infty \leqslant \|A(n-1)\cdots A(m)\|_\infty \otimes \|A(m-1)\cdots A(l)\|_\infty\ ,$$

that is $Z_{l,n} \leqslant Z_{l,m} + Z_{m,n}$.
We have furthermore that $\forall i,j,\forall n$:

$$(A(n-1)\cdots A(0))_{ij} \geqslant A(n-1)_{ii}\cdots A(1)_{ii}A(0)_{ij}\ .$$

If we denote $K_{ij} = E(A(0)_{ij})$ and $K = \min_{ij} K_{ij}$, we conclude that $E(Z_{0,n}) \geqslant K \times n$. We set $e = (e,\dots,e)'$. The sequence $Z_{0,n} = x(n,e)$ verifies the conditions of application of Kingman's sub-additive ergodic theorem, see [97] and also Theorem 10.2.2 for a precise statement. We conclude by remarking that for every finite initial condition x_0, we have :

$$\|x(n,e)\|_\infty \otimes \min_i (x_0)_i \leqslant \|x(n,x_0)\|_\infty \leqslant \|x(n,e)\|_\infty \otimes \max_i (x_0)_i\ .$$

□

Variants and generalizations of Theorem 6.7.1 are proposed in Chapter 9, Theorem 9.3.1 and Chapter 10, Theorem 10.3.5.

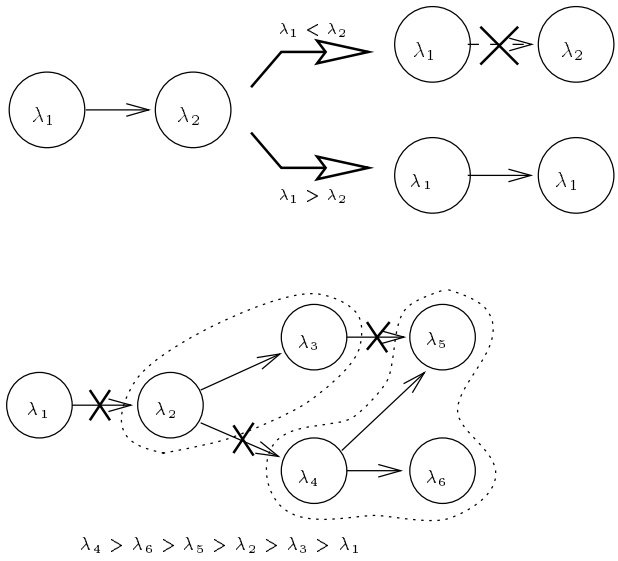


Figure 6.2: Behaviour of open systems.

First order limits for open systems

We assume in this paragraph that matrix $A(0)$ has a fixed structure, see Definition 6.3.5. We decompose the graph of $A(0)$ into its maximal strongly connected subgraphs (s.c.s.). If we replace each s.c.s. by one node, we obtain an associated reduced graph which is acyclic. We associate with each node $\tilde{u}$ of the reduced graph a constant $\lambda_{\tilde{u}}$ which is the Lyapunov exponent of the corresponding s.c.s. in isolation, see Theorem 6.7.1. We denote by ${}^\bullet\tilde{u}$ the set of predecessors of $\tilde{u}$ (including $\tilde{u}$) in the reduced graph. We have :

Theorem 6.7.2 (Baccelli [4]). *Let $\{A(n)\}$ be a stationary and ergodic sequence of matrices. We suppose that $A(0)$ has a fixed structure. We suppose also that $P(A_{ij}(0) = \varepsilon) = 1$ or $\varepsilon < E(A_{ij}(0)) < +\infty, \forall i,j$. Let us consider $i \in \{1,\ldots,k\}$, i belongs to the s.c.s. $\tilde{u}$.*

$$\lim_n \frac{x_i(n,x_0)}{n} = \lim_n E\left(\frac{x_i(n,x_0)}{n}\right) = \bigoplus_{\tilde{v}\in {}^\bullet\tilde{u}} \lambda_{\tilde{v}},\ P-a.s..$$

Intuitively, the dynamic of the system is imposed by the s.c.s. having the smallest throughput (largest cycle time). We propose two illustrative examples in Figure 6.2. In order to get a deeper intuition of this result, one can look at the examples following Theorem 6.7.3.
If the sequence $A(n)$ is i.i.d., some additional results on convergence rates exist. Resing and al [122] prove a Central Limit Theorem in some special cases. Glasserman and Yao [75], Chapter 7, propose results based on a martingale approach. Let us mention also that large deviation results have been proved by Chang [37].

Second order limits for open systems

Matrices $A(n)$ have a fixed structure. In order to simplify the presentation of the results, let us assume that the structure consists of two s.c.s. The general case is completely similar. Up to a

permutation of the coordinates, we have :

$$A(n) = \begin{pmatrix} \tilde{U}(n) & \varepsilon \\ \tilde{B}(n) & \tilde{A}(n) \end{pmatrix} .$$

The block $\tilde{U}$ is a square matrix of size $I \times I$, irreducible. It is interpreted as the input of our system. The block $\tilde{A}$ is a square matrix of size $(k-I) \times (k-I)$, irreducible. The block $\tilde{B}$ is the matrix of the communications between the sources $(\tilde{U})$ and $(\tilde{A})$. We suppose that the block $\tilde{U}$ in isolation has a unique stationary regime (for example if $I = 1$, case of a simple source). We have the following theorem.

Theorem 6.7.3 (Baccelli [4]). *Let u and a be the Lyapunov exponents of $\tilde{U}$ and $\tilde{A}$ respectively (see Theorem 6.7.1). If $a < u$, there is a unique stationary regime for the SRS $\pi(x(n))$, regardless of the initial condition. Convergence to the stationary regime occurs with strong coupling. If $a > u$, then the differences of the form*

$$\delta_{ji}(n, x_0) = x_j(n, x_0) - x_i(n, x_0),\ i = 1, \ldots, I,\ \ j = I+1, \ldots, k ,$$

tend to $+\infty$, $P-a.s.$, for all finite initial condition.

A good way to intuit this result is to consider deterministic matrices. We consider a matrix for which $(u = 1) > (a = e)$.

$$A = \begin{pmatrix} 1 & \varepsilon \\ e & e \end{pmatrix} , \ A^n = \begin{pmatrix} n & \varepsilon \\ n-1 & e \end{pmatrix} .$$

We set $x_0 = (u, v)'$. We have $A^n x_0 = (nu, (n-1)u \oplus v)'$. For n sufficiently large, we have

$$A^n x_0 = nu \otimes \begin{pmatrix} e \\ -1 \end{pmatrix} \Longrightarrow \pi(A^n x_0) = \pi \begin{pmatrix} e \\ -1 \end{pmatrix} .$$

We consider now a case where $(u = e) < (a = 1)$.

$$A = \begin{pmatrix} e & \varepsilon \\ e & 1 \end{pmatrix}, \ A^n = \begin{pmatrix} e & \varepsilon \\ n-1 & n \end{pmatrix} .$$

We have

$$\pi(A^n x_0) = \pi \begin{pmatrix} e \\ (n-1) \oplus (v-u)n \end{pmatrix} .$$

We check that $x_2(n) - x_1(n) = (n-1) \oplus (v-u)n$ tends to $+\infty$ for all finite x_0.

For a stochastic model, the idea remains the same. If $u > a$, the sources which are slower impose their pace. If $u < a$, everything happens asymptotically as if $\tilde{A}$ were in isolation.

Remark In the previous theorem, we need the assumption that $\tilde{U}$ in isolation has a unique stationary regime. But the problem of knowing if $\tilde{U}$ has a unique stationary regime is precisely the one which is going to be addressed in the following. Then, to determine if there is a unique stationary regime for $\pi(x(n))$, we have in fact to use the results of Section 6.7.4 (applied to $\tilde{U}$) together with the comparison of Lyapunov exponents (of $\tilde{U}$ and $\tilde{A}$).

6.7.2 Preliminary results

From now on, we concentrate on second order limits in the closed (i.e $A(n)$ is P-a.s. irreducible) case. The limits are expected to be random variables. We are interested in determining whether the limiting distribution is unique. Furthermore, we want to investigate the type of convergence to the limit. It appears that the notion of coupling convergence is central in our model. In fact, we show that some of the quantities we are interested in are SRS (Def. 6.6.1).

We recall that π is the canonical projection $\mathbb{R}^k \xrightarrow{\pi} \mathbb{PR}^k$ (Def. 6.3.6). It is clear that the recursive equation $x(n+1) = A(n)x(n)$ defines a SRS. It is then easy to show that $\pi(x(n))$ is also a SRS. Indeed, let us consider $x(n)$ and $x'(n)$ such that $\pi(x(n)) = \pi(x'(n))$. We define $x(n+1) = A(n)x(n)$ and $x'(n+1) = A(n)x'(n)$. It is straightforward that $\pi(x(n+1)) = \pi(x'(n+1))$. We write with some abuse of notation that $\{\pi(x(n)),\ n \in \mathbb{N}\}$ verifies the recursive equation "$\pi(x(n+1)) = \pi A(n)\pi(x(n))$"[4].

Proposition 6.7.4. *For $i \in \{1, \ldots, k\}$, we define $z_i(n) = x_i(n) - x_i(n-1)$. The vector $(z_i(n), \pi(x(n)))$ is a SRS.*

Proof. This proposition was proved in [122].

$$\begin{aligned} z_i(n) &= \bigoplus_j [A_{ij}(n-1)x_j(n-1)] - x_i(n-1) \\ &= \bigoplus_j [A_{ij}(n-1) \otimes (x_j(n-1) - x_i(n-1))] \ , \end{aligned}$$

and $(x_j(n-1) - x_i(n-1))$ depends only on $\pi(x(n-1))$. □

More precisely, we have : $\forall i,\ z_i(n) = F_i(\ A(n-1), \pi(x(n-1))\ )$, where F_i is an absolutely continuous function. The sequence $\{A(n)\}$ being stationary, it implies the following corollary.

Corollary 6.7.5. *A sufficient condition for $(z_1(n), \ldots, z_k(n))'$ to converge weakly (resp. in total variation) to a unique invariant distribution, uniformly over initial conditions in $\mathbb{PR}^k$, is that $\pi(x(n))$ has the same property.*

This sufficient condition is not necessary ... as demonstrated by the following deterministic example.

Example 6.7.6. Let us consider

$$A = \begin{pmatrix} e & -1 \\ -1 & e \end{pmatrix} .$$

We have $A^+ = A$, so $u_1 = (e, -1)'$ and $u_2 = (-1, e)'$ are eigenvectors of A. The set

$$\{u_\lambda = \lambda \otimes u_1 \ \oplus\ (1-\lambda) \otimes u_2,\ \lambda \in [0,1]\} \ ,$$

[4] It would be more rigorous to use different notations π and $\tilde{\pi}$ for the canonical projections in $\mathbb{PR}^k$ and $\mathbb{PR}^{k\times k}_{max}$ respectively. Then we would define more formally $\tilde{\pi}(A(n))\pi(x(n)) \equiv \pi(A(n)x(n))$.

is the set of eigenvectors of A, see Theorem 6.4.4. There is a continuum of stationary regimes for $\pi(x(n))$. For example, it is easy to check that for an initial condition u_λ, $\lambda \in [0,1]$, we have:

$$x_1(n, u_\lambda) - x_2(n, u_\lambda) = 2 \times \lambda - 1 \, .$$

But on the other hand, we have a unique stationary regime for $z_i(n)$. As a direct consequence of the equality $A^2 = A$, we have $z_1(n) = z_2(n) = e$, $\forall n \geqslant 2$.
We can also easily build stochastic counter-examples of the same kind.

Remark The variables $\pi(x(n))$ depend only on the sequence $\{\pi(A(n))\}$. Therefore, all the results on $\pi(x(n))$ would still be true under the weaker assumption that only the sequence $\{\pi(A(n))\}$ is stationary and ergodic. But, on the other hand, the variables $z_i(n)$ depend on the sequence $\{A(n)\}$ and not only on $\{\pi(A(n))\}$. Corollary 6.7.5 would not be true under the assumption that $\{\pi(A(n))\}$ only is stationary and ergodic.

In the rest of the paper, we investigate the existence of a stationary regime for the SRS $\pi(x(n))$, i.e. the existence of a finite r.v. $Z : \Omega \to \mathbb{PR}^k$ such that[5]

$$Z \circ \theta = \pi \left(A(0) \pi^{-1}(Z) \right) \, .$$

We are interested by conditions ensuring the uniqueness of the stationary regime and the convergence of $\pi(x(n, x_0))$ toward it, for all $x_0 \in \mathbb{R}^k$. In such cases, we say that the model is *stable*. Two types of convergence will appear, convergence with η-coupling and convergence with coupling. They imply, respectively, weak convergence and total variation convergence as recalled in §6.6.

6.7.3 Stability of discrete models

Let $\{A_l,\ l \in \mathcal{L} \text{ or } l \in \mathbb{N}\}$, be a finite or countable collection of irreducible matrices of size $k \times k$. We suppose that there exists a discrete probability law $\{p_l\}$ such that $A(n, \omega) = A_l$ with probability $p_l > 0$.

Definition 6.7.7 (pattern, 1).
A matrix $\tilde{A}$ is called a pattern of the random sequence $\{A(n), n \in \mathbb{N}\}$ if:

1. $\exists N \mid \tilde{A} = A_{u_{N-1}} \otimes \cdots \otimes A_{u_0}$ *with* $u_0, \ldots, u_{N-1} \in \mathcal{L}$ *(or* $\mathbb{N}$*).*

2. $P(A(N-1) \otimes \cdots \otimes A(0) = \tilde{A}) > 0$.

If the sequence $\{A(n)\}$ is i.i.d. then the second condition is always verified.

Theorem 6.7.8. *The sequence of matrices $\{A(n)\}$ is i.i.d. If there exists a pattern of $\{A(n)\}$ whose critical graph has a unique s.c.s. and whose cyclicity is 1 (***scs1-cyc1** *matrix), then $\{\pi(x(n))\}$ converges with strong coupling to a unique stationary regime. It implies total variation convergence of $\{\pi(x(n))\}$ to its stationary distribution.*

[5] We will write $Z \circ \theta = \pi A(0) Z$ with some abuse of notations.

Proof. Let $C = A_{u_{N-1}} \otimes \cdots \otimes A_{u_0}$ be a scs1-cyc1 pattern. We have, using the cyclicity 1 assumption (Th. 6.4.5),

$$\exists M \mid \forall m \geqslant M,\, C^{m+1} = \lambda C^m \,,$$

where λ is the Lyapunov exponent of C. We conclude that for all initial condition y, $C^{M+1}y = C(C^M y) = \lambda \otimes C^M y$. It means that $C^M y$ is an eigenvector of C. By the assumption on the critical graph of C, there is a unique eigenvector (up to a constant) denoted y_0 (Th. 6.4.4). We have $C^M \otimes y = \mu(y) \otimes y_0$, $\mu(y) \in \mathbb{R}$, or equivalently $\pi(C^M y) = \pi(y_0)$. We define

$$\mathcal{B}_i = \{\omega |\, A(i+MN-1,\omega) \otimes \cdots \otimes A(i+1,\omega) \otimes A(i,\omega) = C^M\} \,.$$

From the **i.i.d.** assumption, it follows that $\mathrm{P}(\mathcal{B}_i) > 0$. On $\mathcal{B}_i$, and for all initial condition y, we have:

$$\begin{aligned} x(i+MN) &= C^M \otimes x(i) \\ &= \mu(x(i)) \otimes y_0 \\ \Longrightarrow \quad \pi(x(i+MN)) &= \pi(y_0) \,. \end{aligned}$$

We check that the sequence $\mathcal{B}_i$ is compatible with the shift, i.e. $\mathcal{B}_i = \mathcal{B}_0 \circ \theta^i$. We conclude that $\mathcal{B}_i$ is a stationary renovating event sequence for the SRS $\pi(x(n))$. We apply Borovkov Theorem (version 6.6.6 for the set $V = \mathbb{PR}^k$, as we have obtained a sequence of renovating events independent of the initial condition) and the uniqueness of the stationary regime follows. □

Remark This theorem is in particular true in the important case where one of the matrices A_l is a scs1-cyc1 matrix.

Example 6.7.9. [Cyclic Jackson Network 2]

We consider a basic Cyclic Jackson Network with k queues and k customers. Such a network can be represented by the (max,+) matrix given in §6.2.2, Equation (6.3). We suppose that the sequence of service times $\{(\sigma_1(n), \ldots, \sigma_k(n)) \,,\, n \in \mathbb{N}\}$ is i.i.d. However the random variables $\sigma_1(n), \ldots, \sigma_k(n)$ need not be independent for a given n. We suppose also that the service times have a discrete support, i.e. can only take a countable number of values. We are in the framework of Theorem 6.7.8. We conclude that a sufficient condition of stability is to find a scs1-cyc1 matrix among the (max,+) matrices corresponding to this network. As a direct application of the result stated in Example 6.4.8, we obtain that a condition of stability is:

$$P(\exists i \mid \sigma_i(n) > \sigma_j(n), \forall j \neq i) > 0 \text{ or } P(\sigma_1(n) = \sigma_2(n) = \cdots = \sigma_k(n)) > 0 \,.$$

We now give a version of Theorem 6.7.8 in the stationary and ergodic case.

Theorem 6.7.10. *The sequence $\{A(n)\}$ is stationary and ergodic. We suppose that there exists a finite pattern $C = A_{u_{N-1}} \otimes \cdots \otimes A_{u_0}$ which is scs1-cyc1 and of rank 1 (see Def. 6.4.7). We suppose that $\mathcal{B} = \{\omega \mid A(N-1)A(N-2)\cdots A(1)A(0) = C\}$ is of strictly positive probability. Then $\{\pi(x(n))\}$ converges with strong coupling to a unique stationary regime.*

Proof. The proof resembles the one of Theorem 6.7.8. As C is of rank 1, we have (see Def. 6.4.7): $C^2 = \lambda \otimes C$, where λ is the Lyapunov exponent of C. We conclude that:

$$\forall y \in \mathbb{R}^k,\ C^2 y = C(Cy) = \lambda C y\ .$$

It implies that Cy is an eigenvector of C. As matrix C is scs1, it has a unique eigenvector y_0, up to a constant. On $\mathcal{B}_i = \mathcal{B} \circ \theta^i$, we have

$$\begin{aligned} \pi(x(i+N)) &= \pi\left(Cx(i)\right) \\ &= \pi(y_0)\ . \end{aligned}$$

We check that the sequence $\mathcal{B}_i$ is compatible with the shift and we apply Borovkov's Theorem 6.6.6. □

Remark If the dependence between matrices is markovian, a sufficient condition to get $P(\mathcal{B}) > 0$ is that $p(A_{u_i}, A_{u_{i+1}}) > 0$, $\forall i = 1, \ldots, N-1$, where $p(.,.)$ is the markovian transition kernel.

Remark The conditions of this theorem are, of course, weaker than the i.i.d. assumption of Theorem 6.7.8. However we made an additional assumption, namely that the pattern C is of rank 1. This assumption cannot be relaxed, as shown by the counter-example 6.7.11.

Example 6.7.11. Let $\Omega = \{\omega_1, \omega_2\}$ be the probability space, P=$\{\frac{1}{2}, \frac{1}{2}\}$ the probability law, and θ the stationary and ergodic shift defined by: $\theta(\omega_1) = \omega_2$ and $\theta(\omega_2) = \omega_1$. We consider

$$A = \begin{pmatrix} 1-\eta & e \\ e & 1 \end{pmatrix},\ B = \begin{pmatrix} 1 & e \\ e & 1-\eta \end{pmatrix},\ e < \eta \ll 1\ .$$

$$\{A(n,\omega_1)\} = A, B, A, B, \ldots \quad \{A(n,\omega_2)\} = B, A, B, A, \ldots\ .$$

Both matrices A and B are scs1-cyc1 patterns of length 1. But patterns which are scs1-cyc1 **and** of rank 1 are for example A^n or B^n for $n > [1/\eta]$. We have, of course, for any $n > [1/\eta]$, $P(\exists N \mid A(N-1)\cdots A(0) = A^n) = P(\exists N \mid A(N-1)\cdots A(0) = B^n) = 0$. Hence the conditions of Theorem 6.7.10 are not verified. In fact, there is a continuum of possible periodic limits. Consider $x_0 = (a,b)'$ with $-1+\eta < a-b < 1-\eta$. Then the limit regime of $\pi(x(n))$ has a state space which is either $\{\pi(a,b)', \pi(a+\eta,b)'\}$ (with probability $\frac{1}{2}$), or $\{\pi(a,b)', \pi(a,b+\eta)'\}$ (with probability $\frac{1}{2}$).
When η becomes arbitrarily small, we can build more complex counter-examples to Theorem 6.7.10 by limiting the number of times that one can get the same matrix A or B in a row. This idea is used in Example 6.8.2. Such models, where there exist patterns which are scs1-cyc1 but no patterns of rank 1, will be called pseudo-periodic.

6.7.4 Stability of general models

In this section, we consider a general model where the coordinates of our matrices have a support which can be discrete, absolutely continuous with respect to Lebesgue measure or a mixture of these two cases.

We need the following definitions, extending the notion of pattern we have been using for finite models. Let M be a deterministic matrix and $\eta > 0$. We denote by $B(M, \eta)$ the open ball of center M and of radius η for the supremum norm of $\mathbb{R}^{k\times k}$. We have $N \in B(M, \eta)$ iff

$$\forall i, j,\ N_{ij} \in]M_{ij} - \eta, M_{ij} + \eta[$$

Definition 6.7.12 (pattern, 2). *Let A be a random matrix. We say that $\tilde{A}$ is a pattern of A if $\tilde{A}$ is a deterministic matrix verifying*

$$\forall \eta > 0,\ P\left\{A \in B(\tilde{A}, \eta)\right\} > 0\ .$$

Equivalently, we can say that $\tilde{A}$ belongs to the support of the random matrix A. It includes the cases where $\tilde{A}$ is an accumulation point (discrete case) or a boundary point (continuous case) of the support.

Definition 6.7.13 (pattern, 3). *Let $\{A(n), n \in \mathbb{N}\}$ be a sequence of random matrices. We say that the deterministic matrix $\tilde{A}$ is a pattern of the sequence $\{A(n)\}$ if*

$$\exists N\ s.t. \forall \eta > 0,\ P\left\{A(N-1) \otimes \cdots \otimes A(0) \in B(\tilde{A}, \eta)\right\} > 0\ .$$

Equivalently, we can say that $\tilde{A}$ is a pattern (Def. 6.7.12) of the random matrix $A(N-1) \otimes \cdots \otimes A(0)$. We say that $\tilde{A}$ is an asymptotic pattern of $\{A(n)\}$ if

$$\forall \eta > 0, \exists N_\eta\ s.t.\ P\left\{\pi(A(N_\eta - 1) \otimes \cdots \otimes A(0)) \in \pi(B(\tilde{A}, \eta))\right\} > 0\ .$$

Remark This definition is coherent with the one given in Definition 6.7.7 for a discrete model. Note that, for convenience reasons, asymptotic patterns are defined in the projective space $\mathbb{PR}^{k\times k}$.

Theorem 6.7.14. *The matrices $A(n)$ are i.i.d. (resp. stationary and ergodic). We suppose that there exists a matrix C which is a pattern of $\{A(n)\}$ (see Def. 6.7.13) and which is scs1-cyc1 (resp. of rank 1). Then the SRS $\{\pi(x(n))\}$ has a unique stationary regime $\{Z \circ \theta^n\}$. The convergence occurs with $\eta-$coupling. It implies weak convergence of $\pi(x(n))$ to its unique stationary distribution.*

Proof. It is done in Appendix, §6.10.2. A stronger version of Theorem 6.7.14, together with a rather different proof, is provided by Theorem 6.8.4. □

Theorem 6.7.15. *The sequence of matrices $\{A(n)\}$ is i.i.d. or stationary and ergodic. We assume that there exists a set $\mathcal{C}$ of matrices such that :*

1. *$\forall C \in \mathcal{C}$, C is a matrix of rank 1.*
2. *$\forall C \in \mathcal{C}$, C is a pattern of $\{A(n)\}$.*
3. *$\exists N \mid P\left(A(N-1) \cdots A(0) \in \mathcal{C}\right) > 0$.*

Then $\{\pi(x(n))\}$ converges with strong coupling to a unique stationary regime.

The conditions of Theorem 6.7.15 are stronger than the ones of Theorem 6.7.14 as we require the patterns of rank 1 to be of positive probability. On the other hand, we obtain a stronger type of convergence.

Proof. Let us define $\mathcal{B} = \{\omega \mid A(N-1)A(N-2)\cdots A(1)A(0) \in \mathcal{C}\}$ and $\mathcal{B}_i = \mathcal{B} \circ \theta^i$. Using that the matrices $C \in \mathcal{C}$ are of rank 1, we obtain that, on the event $\mathcal{B}_i$, $\pi(x(i+N))$ is independent of the value of $\pi(x(i))$. It implies that $\{\mathcal{B}_n, n \in \mathbb{N}\}$ is a stationary sequence of renovating events. The result follows. □

Remark Theorems 6.7.8 to 6.7.15 do not require any aperiodicity (Def. 6.3.4) assumption on the matrices $A(n)$. However, the pattern C whose existence is essential in all of these theorems is aperiodic. The condition "scs1-cyc1" implies aperiodicity.

Example 6.7.16. [Cyclic Jackson Network 3]

We consider the same i.i.d. model as in Example 6.7.9. However, the distributions of the service times are now general. We obtain, by using Theorems 6.7.14 and 6.7.15, the stability under the condition:
The support of the random vector $(\sigma_1(n), \ldots, \sigma_k(n))$ contains at least one point such that:

$$\exists i \mid \sigma_i(n) > \sigma_j(n), \forall j \neq i \text{ or such that } \sigma_1(n) = \sigma_2(n) = \cdots = \sigma_k(n) \; .$$

If the previous condition occurs with strictly positive probability, we obtain total variation convergence. Otherwise, we obtain weak convergence. Here is a case with only weak convergence. We consider an i.i.d. CJN with three queues and three customers. We assume that $\sigma_1 = \sigma_2 = 1$ and σ_3 is uniformly distributed over $[0, 1]$.

6.8 Converse Theorems

We are going to prove converses of Theorems 6.7.8, 6.7.10, 6.7.14 and 6.7.15. We will consider successively finite and general models of type: "$x(n+1) = A(n)x(n)$" where the matrices are of size $k \times k$ and are $P-a.s$ irreducible. We will, moreover, always suppose that there exists, with positive probability, a pattern whose projective diameter (Def. 6.3.9) is finite, i.e.:

$$\exists n \mid P\{\mathtt{D}(\, A(n)\cdots A(0)\,) < +\infty\} > 0 \; . \tag{6.6}$$

It implies $\lim_n P\{\mathtt{D}(\, A(n)\cdots A(0)\,) < +\infty\} = 1$, see the proof of Lemma 6.10.1. This condition is very weak. In the i.i.d. case, it is enough that there exists a pattern which is irreducible and aperiodic. We comment further on this condition in Section 6.9.1.

6.8.1 Finite models in $\mathbb{Q}_{max}^{k\times k}$

We consider a finite model: "$x(n+1) = A(n)x(n)$", with $A(n) \in \{A_l,\ l \in \mathcal{L} = \{1, \ldots, L\}\}$. We assume that the matrices are irreducible. We assume also that the matrices $A_l, l \in \mathcal{L}$, belong to $\mathbb{Q}_{max}^{k\times k}$, i.e. that their coordinates are rational.

Theorem 6.8.1. *The sequence of matrices $\{A(n)\}$ is i.i.d. or stationary and ergodic. When there is a unique stationary regime, convergence to this regime occurs with strong coupling. A necessary and sufficient condition for the model to have a unique stationary regime is that there exists a matrix C verifying*

1. *C is a matrix of rank 1 (Def. 6.4.7).*
2. *C is a pattern of $\{A(n)\}$ (Def. 6.7.13).*

Proof. It is given in Appendix, §6.10.3. □

Theorem 6.8.1 is not true in general when the matrices $A_l, l \in \mathcal{L}$, belong to $\mathbb{R}_{max}^{k\times k}$, see the following counter-example.

Example 6.8.2. We consider the matrices

$$A = \begin{pmatrix} e & -1 \\ -1 & -\eta \end{pmatrix}, \; B = \begin{pmatrix} -\eta' & -1 \\ -1 & e \end{pmatrix},$$

where $0 < \eta, \eta' \ll 1$ and η, η' are not co-rational, i.e. $\eta/\eta' \notin \mathbb{Q}$.

Let $u = (u_1, u_2)' \in \mathbb{R}^2$, we set $\psi(u) = u_2 - u_1$. We identify $\mathbb{PR}^2$ and $\mathbb{R}$ using the function $\psi \circ \pi^{-1}$. The matrices A and B are scs1-cyc1. Their respective and unique eigenvectors are $\psi(e_1) = -1$ and $\psi(e_2) = 1$. For a vector $u = (u_1, u_2)'$ such that $\psi(u) \in [-1, 1]$, we have

$$\psi(Au) = \max(\psi(u) - \eta, -1), \;\; \psi(Bu) = \min(\psi(u) + \eta', 1) \; . \tag{6.7}$$

We consider a Markov chain defined on the set $\psi^{-1}[-1, 1] \subset \mathbb{R}^2$. The transition probabilities are

- For u such that $\psi(u) \in]-1+\eta, 1-\eta'[$, $p(u, Au) = 1/2$, $p(u, Bu) = 1/2$.
- For u such that $\psi(u) \in [1-\eta', 1]$, $p(u, Au) = 1$.
- For u such that $\psi(u) \in [-1, -1+\eta]$, $p(u, Bu) = 1$.

The behaviour of the Markov chain is illustrated in Figure 6.3.

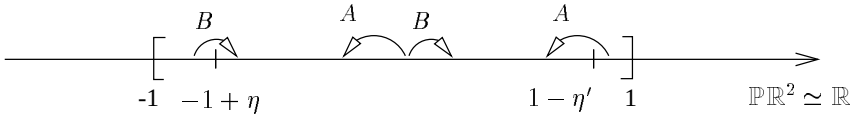


Figure 6.3: Markov chain $\psi(X(n))$ on $\mathbb{R}$.

Let $X(n)$ be a realization of the Markov chain. It is easy to check that this Markov chain is aperiodic. Under the assumption $\eta/\eta' \notin \mathbb{Q}$, one can prove using classical arguments that the set $\{\psi(X(n)), n \in \mathbb{N}\}$ is P-a.s. dense in $[-1, 1]$. It implies that the Markov chain is ν-irreducible where ν is the Lebesgue measure on $\psi^{-1}[-1, 1]$. Hence there exists a unique stationary distribution Q for the Markov chain. It verifies $Q(\mathcal{A}) > 0$ for all event $\mathcal{A}$ such that

$\nu(\mathcal{A}) > 0$. For a complete presentation of Markov chains on continuous state spaces, see Meyn & Tweedie [109].
Let us consider a stationary realization $X(n)$ of the Markov chain (i.e. $\forall n$, $P\{X(n) \in .\} = Q(.)$). We define

$$A(n,\omega) = \begin{cases} A & \text{if } X(n+1,\omega) = AX(n,\omega), \\ B & \text{if } X(n+1,\omega) = BX(n,\omega). \end{cases}$$

As $X(n)$ is stationary, it follows that $\{A(n)\}$ is a stationary and ergodic sequence.

Let us consider the stationary-ergodic finite model "$x(n+1) = A(n)x(n)$" and $x(0) = x_0 \in \mathbb{R}^2$ is non-random. Note that $\{x(n)\}$ is not a Markov chain anymore.
Let us consider a pattern $C = A_{n-1} \cdots A_0$ of $\{A(n)\}$, i.e. $P\{A(n-1) \cdots A(0) = A_{n-1} \cdots A_0\} > 0$. Let $x_0, \ldots, x_n$ be a corresponding path for the Markov chain $X(n)$, i.e.

$$x_0 \in \psi^{-1}]-1,1[, \qquad \begin{array}{l} x_1 = A_0 x_0, \ldots, x_n = A_{n-1} x_{n-1}, \\ \text{and } P\{X(n) = x_n, \ldots, X(1) = x_1 \mid \ X(0) = x_0\} > 0\,. \end{array}$$

Let us denote by $\underline{c}$ and $\overline{c}$ the minimal distances between $x_p, p \leqslant n$ and the extremal points of $\psi^{-1}[-1,1]$.

$$\underline{c} = \min_{p \leqslant n} (\psi(x_p) + 1),\ \overline{c} = \min_{p \leqslant n} (1 - \psi(x_p))\ .$$

It follows from (6.7) that

$$\psi(\mathrm{Im}(A_{n-1} \cdots A_0)) = [x_n - \underline{c}, x_n + \overline{c}]\,, \tag{6.8}$$

where $\mathrm{Im}(A) = \{Au,\ u \in \mathbb{R}^k\}$. From the definition of the Markov chain $X(n)$, it follows that $\underline{c} > 0, \overline{c} > 0$. We conclude that $A_{n-1} \cdots A_0$ is not a rank 1 matrix. There exists no finite pattern of rank 1 for $\{A(n)\}$.

On the other hand, let us prove that there exists asymptotic patterns of rank 1 for $\{A(n)\}$. We define $\underline{c}(n) = \min_{p \leqslant n}(\psi(X(p)) + 1)$ and $\overline{c}(n) = \min_{p \leqslant n}(1 - \psi(X(p)))$. As $\{\psi(X(n)), n \in \mathbb{N}\}$ is dense in $[-1,1]$, we obtain that $\underline{c}(n) \to 0$ and $\overline{c}(n) \to 0$. Using (6.8), we obtain that $\mathtt{D}(A(n) \cdots A(0)) \to 0$, P-a.s. We conclude following the lines of Theorem 6.8.4, §6.10.4. There is a unique stationary regime for the model. For an arbitrary initial condition, we have η-coupling (weak convergence) with this stationary regime.

To summarize, we have exhibited a finite model with a unique stationary regime and no coupling convergence. This type of behaviour is closely related to the non-finiteness of the projective semigroup $\pi < A, B >$ (see Def. 6.5.1 for a definition and Chapter 3, §3.7 for more insights).

6.8.2 General models

We consider a general model of type "$x(n+1) = A(n)x(n)$". Stability no longer implies coupling in finite time. It was illustrated by Example 6.8.2. Here is another example, for an i.i.d. model.

Example 6.8.3.

$$A(n) = \begin{pmatrix} U(n) & e \\ e & U(n) \end{pmatrix},$$

where $U(n)$ are i.i.d. random variables of uniform distribution over $[0,1]$. There is a unique stationary regime for $\pi(x(n))$ which is $\pi(e,e)'$. We denote by $d(.,.)$ the projective distance. For an initial condition $(y,e)'$ with $y \geqslant 1$, we have $d\left(x(n);(e,e)'\right) = \min_{p \leqslant n} U(p)$. Thus convergence to $\pi(e,e)'$ occurs only asymptotically. There is no coupling but only η-coupling with the unique stationary regime.

We can show the following results.

Theorem 6.8.4. *The sequence of matrices $\{A(n)\}$ is i.i.d. or stationary and ergodic. The necessary and sufficient condition for the model to converge with η-coupling to a unique stationary regime is the existence of an asymptotic pattern C of $\{A(n)\}$ of rank 1 (Def. 6.7.13).*

Proof. It is given in Appendix, §6.10.4. □

Theorem 6.8.5. *The sequence of matrices $\{A(n)\}$ is i.i.d. or stationary and ergodic. The necessary and sufficient conditions for the model to converge with coupling to a unique stationary regime are :*
There exists a set $\mathcal{C}$ of matrices such that :

1. $\forall C \in \mathcal{C}$, C *is a matrix of rank 1.*
2. $\forall C \in \mathcal{C}$, C *is a pattern of* $\{A(n)\}$.
3. $\exists N \mid P\left(A(N-1)\cdots A(0) \in \mathcal{C}\right) > 0$.

We can say equivalently that we must have patterns of rank 1 but with strictly positive probability.

Proof. We have already proved the sufficient part (Th. 6.7.15). We prove the necessary part of the theorem in Appendix, §6.10.5. □

Remark Convergence with $\eta-$coupling appears as a limiting case of coupling in finite time. In a discrete model, we will have only η-coupling when the set $\mathcal{C}$ of scs1-cyc1 patterns is non empty but is of probability 0. It means that the scs1-cyc1 patterns are only accumulation points of the support. In a general model, we will have only η-coupling when the scs1-cyc1 patterns are isolated points of the support (which implies that they are boundary points of the support).

Example 6.8.6. To illustrate the previous remark, let us continue the analysis of Example 6.8.3. There is only one matrix (in the projective space $\mathbb{PR}_{max}^{k\times k}$) verifying the first two conditions of Theorem 6.8.5. It is the matrix

$$\pi(C) = \pi \begin{pmatrix} e & e \\ e & e \end{pmatrix},$$

But condition 3. of Th. 6.8.5 is not verified as $\forall N, P\left(\pi(A(N-1)\cdots A(0)) = \pi(C)\right) = 0$.
Let us consider a slightly modified sequence of matrices $\{\tilde{A}(n)\}$ where the diagonal elements are two random variables $U(n)$ and $U'(n)$ defined on $[0,1]$ and such that

$$P\left\{U(n) > U'(n)\right\} > 0 \text{ or } P\left\{U'(n) > U(n)\right\} > 0\,.$$

Now, we have scs1-cyc1 patterns with strictly positive probability and there is coupling in finite time with the unique stationary regime.

Example 6.8.7. [Cyclic Jackson Network 4]

We consider the model of Example 6.7.16. The condition

$$P\left(\{\exists i \mid \sigma_i(n) > \sigma_j(n), \forall j \neq i\} \cup \{\sigma_1(n) = \cdots = \sigma_k(n)\}\right) > 0\,.$$

is necessary and sufficient for strong coupling convergence to a unique stationary regime. For i.i.d. Cyclic Jackson Networks, the sufficient condition $P\left(\{\exists i \mid \sigma_i(n) > \sigma_j(n), \forall j \neq i\}\right) > 0$ was obtained in [94]. The method of proof was completely different, see the remarks at the end of Section 6.2.2.

In Figure 6.4, at the end of the chapter, we propose a diagram summarizing the different results proved.

6.9 Complementary Results

6.9.1 Without irreducibility

We have supposed from the beginning that the matrices $\{A(n)\}$ were irreducibles. The relaxation of the irreducibility assumption is very important in terms of modeling power. It enables us to consider, for example, the task graphs with random precedences introduced in Section 6.2.1.

Let us come back to the places where this assumption is used. First of all, it is used in Prop. 6.3.8. But in fact, the only point we need in order to prove this proposition is: "$\forall u \in \mathbb{R}^k,\ Au \in \mathbb{R}^k$", i.e. if u has only non-ε coordinates then Au has the same property. So the only assumption we need on the matrices $\{A(n)\}$ is:

I $\quad \forall i,\ P\left\{\exists j \text{ s.t. } A_{ij}(0) > \varepsilon\right\} = 1\,.$

Secondly, the irreducibility is essential for the $\mathbb{R}_{max}$ spectral theory of Section 6.4. A reducible matrix $A \in \mathbb{R}^{k\times k}$ may have several eigenvalues. Definition 6.4.3 and Theorem 6.4.4 have to be reinterpreted by replacing the unique eigenvalue by the maximal eigenvalue. Theorem 6.4.5 is not true anymore. But as far as the direct theorems (6.7.8, 6.7.10, 6.7.14 and 6.7.15) are concerned, we use results of the $\mathbb{R}_{max}$ spectral theory only for the pattern C whose existence is critical for the proofs. They are still valid, then, if we state that condition **I** is verified and that there exists a pattern C which is scs1-cyc1 **and** irreducible.

Dropping the irreducibility assumption does not influence the converse results. More precisely, the proofs of Theorems 6.8.1, 6.8.4 and 6.8.5 are still valid. Only two points need to be verified:

I $\quad \forall i,\ P\{\exists j \text{ s.t. } A_{ij}(0) > \varepsilon\} = 1\ .$

II $\quad \exists n \mid P\{\mathtt{D}(\ A(n)\cdots A(0)\ ) < +\infty\} > 0.$

Of course, irreducibility $P-a.s.$ is not necessary to ensure that these conditions hold. We conclude that we can state our converse results under the previous two minimal assumptions.

Let us discuss condition **II** a little further. First of all, we propose a counter-example showing that without **II**, the uniqueness of the stationary regime does not imply the existence of a rank 1 pattern.

Example 6.9.1. Let $\Omega = \{\omega_1, \omega_2\}$ be the probability space, $P = \{\frac{1}{2}, \frac{1}{2}\}$ the probability law, and θ the ergodic shift defined by: $\theta(\omega_1) = \omega_2$ and $\theta(\omega_2) = \omega_1$. We consider

$$A = \begin{pmatrix} \varepsilon & e \\ e & 1 \end{pmatrix},\ B = \begin{pmatrix} e & e \\ 1 & \varepsilon \end{pmatrix}\ .$$

$$\{A(n,\omega_1)\} = A, B, A, B, \ldots \quad \{A(n,\omega_2)\} = B, A, B, A, \ldots\ .$$

All patterns have an infinite projective diameter. Therefore, condition **II** is not verified. Nevertheless, there is coupling in finite time with a unique periodic regime. More precisely, there is coupling of $\pi(x(n,u))$ to the periodic regime $\{\pi(e,e)', \pi(e,1)'\}$ and coupling occurs for $n > 2\times[\bigoplus(u_1-u_2, u_2-u_1)]$. We conclude that there is coupling in finite time to a unique stationary regime but no rank 1 pattern. Without condition **II**, Theorem 6.8.1 is not true anymore.

Another class of systems where condition **II** is not verified is the class of open systems studied by Baccelli in [4]. The results for this type of systems have been recalled in §6.7.1. In this case also, Theorem 6.8.1 fails to be true. In such models, there are no patterns which are scs1-cyc1 and irreducible (matrices are non-irreducible with probability 1 !), even when there is a unique stationary regime. In fact, we cannot exhibit a type of pattern which would decide the uniqueness of the stationary regimes. The good criterion is the comparison between Lyapunov exponents, see Theorem 6.7.3. The computation of such exponents involves the whole structure of the stochastic matrices $\{A(n)\}$, and not only an extracted pattern.

For the class of open systems, it is not even true that coupling of the trajectories mean the existence of a stationary regime, see Example 6.9.2. Hence Lemma 6.10.1 fails to be true.

Example 6.9.2. Let us consider the i.i.d. model "$x(n+1) = A(n)x(n)$" with

$$A(n) = \begin{pmatrix} e & \varepsilon \\ e & a(n) \end{pmatrix},$$

and the random variables $a(n)$ are such that $P\{a(n) = -1\} = P\{a(n) = 1\} = 1/2$. We identify $\mathbb{PR}^2$ and $\mathbb{R}$ in the following way $\pi(a,b)' \simeq b - a$. For an initial condition $u = (u_1, u_2)'$, we have $x(n+1) = (u_1, \max(x(n)_1 + a(n), u_1)'$ which implies

$$\pi(x(n+1)) = \max(\pi(x(n)) + a(n), 0)\ .$$

This is a classical birth and death process with state space $\mathbb{N}$. Two trajectories which cross each other get coupled as the only jumps allowed are -1 and 1. We conclude that there is coupling in finite time of the trajectories of $\pi(x(n))$. However the process is recurrent null as a consequence of $E(a(0)) = 0$. Hence there exists no stationary distribution.

Condition **II** is weak and will be verified in most cases. For a discrete i.i.d. model, for example, it is sufficient to have one pattern of finite length $A_{u_N} \cdots A_{u_1}$ which is irreducible and aperiodic to verify it. For a general i.i.d. model, it is sufficient to have $P\{ A(0) \text{ irreducible and aperiodic} \} > 0$. In a stationary and ergodic framework, condition **II** is a little bit stronger, as shown by Example 6.9.1 where $P\{ A(0) \text{ irreducible and aperiodic} \} = 1$ and where condition **II** is not verified.

Remark For a general model which does not verify condition **II**, we decompose the model into its maximal sub-models verifying it. Then the complete analysis of the system boils down to an analysis of the sub-models (using the results of Section 6.7.4) and of their interactions (using Theorem 6.7.3 and its generalizations, see [4]).

6.9.2 Boundedness

We have seen that we do not need irreducibility and aperiodicity assumptions in order to get our results. Anyway, there are more precise results when we make these assumptions.

Proposition 6.9.3. *We consider a finite model verifying the assumptions of Theorem 6.8.1. If we assume furthermore that all the matrices $A_{u_1} A_{u_2} \cdots A_{u_n}, u_i \in \{1, \ldots, L\}$ are aperiodic, then the unique stationary distribution has a bounded state space.*

Proof. The existence and uniqueness of the stationary regime is a consequence of Theorem 6.7.8. We define: $\mathrm{Im}(A_l) = \left\{ v \in \mathbb{R}^k \mid \exists u \in \mathbb{R}^k,\ v = A_l u \right\}$. We recall that a matrix has a bounded image if and only if all its coordinates are different from ε. Let us consider $S =< A_1, \ldots, A_L >$ the semigroup generated by the matrices $A_1, \ldots, A_L$ (Def. 6.5.1). It follows from Proposition 6.5.3 that the semigroup S is primitive. Let N be such that

$$\forall n \geqslant N,\ u_1, \ldots, u_n \in \{1, \ldots, L\},\ \forall i, j,\ (A_{u_1} A_{u_2} \cdots A_{u_n})_{ij} > \varepsilon\ .$$

We define $\Pi = \cup(\mathrm{Im}(A_{u_1} \cdots A_{u_N}))$, where the union is taken over all the products of length N. It follows from the primitivity that Π is bounded. We have that $\forall x_0 \in \mathbb{R}^k,\ \pi(x(n, x_0)) \in \Pi$ for $n \geqslant N$. It implies that the support of the stationary measure is included in Π. □

Proposition 6.9.3 is not true without the aperiodicity assumption, as is shown by Examples 6.9.4 and 6.9.5.

Example 6.9.4. Let us consider an i.i.d. model with 2 matrices and verifying conditions **I** and **II**. Let $p_1 = P(A(0) = A_1)$ and $p_2 = 1 - p_1 = P(A(0) = A_2)$.

$$A_1 = \begin{pmatrix} e & \varepsilon \\ \varepsilon & 1 \end{pmatrix},\ A_2 = \begin{pmatrix} 1 & e \\ e & 1 \end{pmatrix}\ .$$

We identify $\mathbb{PR}^2$ and $\mathbb{R}$ in the following way: $\pi(a, b)' \simeq b - a$. We consider an initial condition $u = (u_1, u_2)'$ with $u_2 - u_1 \geqslant 1$. One verifies that $\pi(x(n, x_0))$ is a Markov chain on $\mathbb{N} \setminus \{0\}$ whose transition probabilities are

$$p(i, i+1) = p_1,\ p(i, 1) = 1 - p_1,\ \ \forall i \in \mathbb{N} \setminus \{0\}\ .$$

This is a classical recurrent positive Markov chain whose stationary distribution is:

$$\mu(1) = 1 - p_1,\ \mu(n) = (1-p_1)p_1^{n-1},\ \forall n \geqslant 2\ .$$

The state space is unbounded. The scs1-cyc1 pattern proving the stability is for example $C = A_1A_2$.

There exist also examples of models where all the matrices are irreducible and where the state space of the unique stationary regime is unbounded.

Example 6.9.5. We consider:

$$B_1 = \begin{pmatrix} \varepsilon & 1/2 \\ e & \varepsilon \end{pmatrix},\ B_2 = \begin{pmatrix} \varepsilon & e \\ 1/2 & \varepsilon \end{pmatrix},\ B_3 = \begin{pmatrix} 1 & e \\ e & 1 \end{pmatrix}\ .$$

A possible pattern to prove the uniqueness of the stationary regime is $C = B_1B_2B_3$. We show that the state space is unbounded by remarking that $B_1B_2 = A_1$ and $B_3 = A_2$, where A_1 and A_2 are defined in Example 6.9.4.

6.10 Appendix

6.10.1 Loynes scheme

Lemma 6.10.1 is going to be used in several of the forthcoming proofs. Under an assumption of η-coupling of the trajectories, we build a stationary regime using a Loynes' type construction.

Lemma 6.10.1. *We consider a general model "$x(n+1) = A(n)x(n)$" (see §6.7.4). The sequence $\{A(n)\}$ is stationary and ergodic. We assume that there exists N such that*

$$P\left\{\mathtt{D}(\ A(N)\cdots A(0)\ ) < +\infty\right\} > 0$$

We assume also that $\forall x_0, y_0 \in \mathbb{R}^k$, $d(x(n,x_0), x(n,y_0)) \to 0$, P a.s. (η-coupling of the trajectories). Then there exists a r.v. $Z : \Omega \to \mathbb{P}\mathbb{R}^k$ verifying $Z \circ \theta = \pi(A(0))Z$. The sequence $\{Z \circ \theta^n\}$ is the unique stationary regime of the model.

Proof. We are going to show that the sequence $\{\pi(A(-1)\cdots A(-n)u),\ n \in \mathbb{N}\}$, $u \in \mathbb{R}^k$, has a simple limit in $\mathbb{P}\mathbb{R}^k$. The argument is an analog of the famous backward scheme proposed by Loynes in [103] for G/G/1 queues.
We want to show that $\mathtt{D}(\ A(n)\cdots A(0)\ ) \overset{n}{\to} 0$, $P-a.s.$ It is easy to see that the event

$$\mathcal{A} = \{\omega \mid \exists N,\ \mathtt{D}(\ A(N)\cdots A(0)\ ) < +\infty\} = \{\omega \mid \exists N, \forall n \geqslant N,\ \mathtt{D}(\ A(n)\cdots A(0)\ ) < +\infty\}$$

is invariant by the translation shift. Then by the ergodic Lemma 10.2.1, it is of probability 0 or 1. We have made the assumption that $P(\mathcal{A}) > 0$, hence $P(\mathcal{A}) = 1$.
Using the stationarity of the sequence $\{A(n), n \in \mathbb{Z}\}$, we have that $\exists N(\omega)$ such that

$$\mathtt{D}\left(\ A(-1)\cdots A(-N)\ \right) < +\infty\ .$$

Then we can define the projective image of $A(-1)\cdots A(-N)$ which is a bounded subset of $\mathbb{PR}^k$ and that we denote by Π. The boundedness implies that

$$c = \max_{v\in\Pi} d(e,v) < +\infty \ ,$$

where $e = (e,\ldots,e)'$. Let us define the vectors

$$c_1 = (c,e,\ldots,e)', c_2 = (e,c,e,\ldots,e)', \ldots, c_k = (e,\ldots,e,c)' \ . \tag{6.9}$$

It is immediate that Π is included in the convex hull of these vectors, i.e.

$$\Pi \subset \{\pi(\alpha_1 \otimes c_1 \oplus \alpha_2 \otimes c_2 \cdots \oplus \alpha_k \otimes c_k),\ \alpha_i \in \mathbb{R}\} \ .$$

In the (max,+) algebra, we have the following property, for all $A \in \mathbb{R}^{k\times k}_{max}, u, v \in \mathbb{R}^k_{max}$, $A(u\oplus v) = Au \oplus Av$. It implies

$$\forall x \in \Pi,\ \pi(Ax) \in \{\pi(\alpha_1 \otimes Ac_1 \oplus \cdots \oplus \alpha_k \otimes Ac_k),\ \alpha_i \in \mathbb{R}\} \ . \tag{6.10}$$

We fix $\eta > 0$. Using the $\eta-$coupling assumption, we have that the random variable $N'(\omega)$ is $P-a.s.$ finite, where N' is defined by:

$$N' = \inf\{n \mid d(x(n,c_i),x(n,c_j)) \leqslant \eta,\ \forall i,j\} \ .$$

As both N and N' are $P-a.s.$ finite, we have

$$\forall \delta > 0,\ \exists L, L' \ :\ \ P\{N \leqslant L, N' \leqslant L'\} \geqslant 1-\delta \ .$$

As a direct consequence of (6.10), we have on the event $\{N \leqslant L, N' \leqslant L'\}$:

$$\mathtt{D}\left(A(L'-1)\otimes\cdots\otimes A(0)A(-1)\otimes\cdots\otimes A(-L)\right) \leqslant \eta \ .$$

We deduce, using the stationarity of $\{A(n)\}$, that

$$\begin{aligned} P\left\{\mathtt{D}(A(-1)\cdots A(-L-L')) \leqslant \eta\right\} &= P\left\{\mathtt{D}(A(L'-1)\cdots A(-L)) \leqslant \eta\right\} \\ &\geqslant 1-\delta \ . \end{aligned}$$

It implies that the random variables $\mathtt{D}(A(-1)\cdots A(-n))$ converge in probability to 0. But as $\mathtt{D}(A(-1)\cdots A(-n))$ is pathwise decreasing, the convergence occurs also $P-a.s.$

We have in particular, for all $u \in \mathbb{R}^k$, $d\left(A(-1)\cdots A(-n)u, A(-1)\cdots A(-n-p)u\right) \longrightarrow 0$, $P-a.s.$ It implies that $\{\pi(A(-1)\cdots A(-n)u)\}$ is a Cauchy sequence which converges. The limit does not depend on u. We denote it by Z. We have :

$$\begin{aligned} Z\circ\theta &= \lim_n \pi(A(0)A(-1)\cdots A(-n)u) \\ &= \pi A(0) \lim_n \pi(A(-1)\cdots A(-n)u) = \pi(A(0))Z \ . \end{aligned}$$

The sequence $\{Z \circ \theta^n\}$ is a stationary regime. Let us prove it is the unique one. We want to prove that

$$\forall x_0 \in \mathbb{R}^k, \;\; d\left(x(n, x_0), Z \circ \theta^n\right) \stackrel{n \to +\infty}{\longrightarrow} 0 \;, \; P - a.s. \tag{6.11}$$

As Z is $P-a.s.$ finite, for all $\eta > 0$, there exists a compact $K \in \mathbb{PR}^k$ such that $P\{Z \in K\} > 1-\eta$. We proceed as above (Equation (6.9)) in order to define vectors $c_i, i = 1, \ldots, k$ such that

$$K \subset \{\pi(\alpha_1 \otimes c_1 \oplus \alpha_2 \otimes c_2 \cdots \oplus \alpha_k \otimes c_k), \; \alpha_i \in \mathbb{R}\} \;.$$

We have

$$Z \in K \Rightarrow Z \circ \theta^p \in \{\pi(\alpha_1[A(p-1)\cdots A(0)c_1] \oplus \cdots \oplus \alpha_k[A(p-1)\cdots A(0)c_k]), \; \alpha_i \in \mathbb{R}\} .$$

Using the $\eta-$coupling of trajectories, we also have

$$\forall c_i, \; d(x(n, x_0), x(n, c_i)) \to 0 \;.$$

We conclude easily that there is $\eta-$coupling of $\{\pi x(n, x_0)\}$ and $\{Z \circ \theta^n\}$. We can apply Proposition 6.6.10. There is weak convergence of $\{\pi x(n, x_0)\}$ to the distribution of Z and relation (6.11) establishes the a.s. convergence of $\{\pi x(n, x_0)\}$ to $Z \circ \theta^n$. As a direct consequence, $\{Z \circ \theta^n\}$ is the unique stationary regime.

In fact it is not necessary to use Proposition 6.6.10. The backward scheme gives us the following result.

$$\forall x_0 \in \mathbb{R}^k, \; \pi(\, A(-1)A(-2)\cdots A(-n)x_0 \,) \stackrel{n \to +\infty}{\longrightarrow} Z \;, \; P - a.s.$$

We consider a function $f : \mathbb{PR}^k \to \mathbb{R}$, continuous and bounded. We have

$$\begin{aligned} E\left(\, f(\pi x(n, x_0)) \,\right) &= E(\, f(\pi A(n-1)\cdots A(0)x_0) \,) \\ &= E(\, f(\pi A(-1)\cdots A(-n)x_0) \,) \stackrel{n}{\longrightarrow} E(\, f(Z) \,) \;, \; P - a.s. \;, \end{aligned}$$

using Lebesgue dominated convergence theorem (f is bounded). It proves weak convergence. We conclude in the same manner. The introduction of the notion of η-coupling is useful to show the continuity with the finite model where there is coupling. Furthermore, if we assume that there exists a solution to $Z \circ \theta = \pi(A(0))Z$, Prop. 6.6.10 enables us to prove the uniqueness of the stationary regime without needing a backward scheme. □

Remark Without the assumption $P\{\mathtt{D}(\, A(N)\cdots A(0) \,) < +\infty\} > 0$, it is not always true that η-coupling of the trajectories implies the existence of a stationary regime, see Example 6.9.2.

6.10.2 Proof of Theorem 6.7.14

The general idea consists in using Theorems 6.7.8 and 6.7.10 after having discretized the matrices $A(n)$. This discretization goes in the following way. We consider a_{ij}, $i, j = 1, \ldots, k$. We define A_0 by $(A_0)_{ij} = a_{ij}$. In general, we have: $P(A(0, \omega) = A_0) = 0$. We consider $\eta > 0$ fixed. We

define A^η, the discretization of step η and of skeleton A_0, in the following way. For i,j fixed, we define:

$$A^\eta_{ij}(0,\omega) = (a_{ij} + 2n\eta) \text{ if } A_{ij}(0,\omega) \in]a_{ij} + (2n-1)\eta, a_{ij} + (2n+1)\eta],\ n \in \mathbb{Z}\,.$$

We check easily that we have $\|A(n,\omega) - A^\eta(n,\omega)\|_\infty \leqslant 2\eta$ with probability 1. Furthermore, the random variables $A^\eta(0)$ converge to $A(0)$ a.s. as $\eta \to 0$.

In the whole proof, $d(.,.)$ represents the projective distance as defined in Def. 6.3.7. Let N be such that C is a rank 1 pattern of $A(N-1) \otimes \cdots \otimes A(0)$. We can write C in the following form: $C = C_{N-1} \otimes \cdots \otimes C_0$ where C_i is a pattern of $A(i)$ (i.e. of $A(0)$), $\forall i = 0, \ldots, N-1$. We define a decreasing sequence $\eta_i > 0$ such that: $8N\eta_i \leqslant \frac{1}{i}$. The N matrices $C_0, \ldots, C_{N-1}$ define N possible skeletons for discretizing matrices $\{A(n)\}$. More precisely, for $l \in \{0, \ldots, N-1\}$, we can define the intervals: $](C_l)_{ij} + (2p-1)\eta, (C_l)_{ij} + (2p+1)\eta]$, $p \in \mathbb{Z}$. By realizing all possible intersections between these intervals, we define a new countable set of disjoint intervals, whose union is $\mathbb{R}$. The discretization of $A(n)$ will be done with respect to this new set. It is straightforward to prove that

$$\forall u, v \in \mathbb{R}^k,\ \ |\, d(A(0)u, A(0)v) - d(A^\eta(0)u, A^\eta(0)v)\,| \leqslant 8\eta\,. \tag{6.12}$$

We are now going to prove that $\forall u, v$, we have

$$d(A(n)\cdots A(0)u, A(n)\cdots A(0)v) \overset{n\to+\infty}{\longrightarrow} 0\ \ P-a.s.$$

We define the events $\mathcal{B}^i_n$ in the following way:

$$\mathcal{B}^i_0 = \{\omega \mid A^{\eta_i}(N-1,\omega) \otimes \cdots \otimes A^{\eta_i}(0,\omega) = C\}\,,\ \mathcal{B}^i_n = \mathcal{B}^i_0 \circ \theta^n\ .$$

We fix i. The event $\mathcal{A} = \left\{\omega \mid \bigcup_{n=0}^\infty \mathcal{B}^i_n\right\}$ is such that $\theta(\mathcal{A}) \subset \mathcal{A}$. By the ergodic lemma, we deduce that it is of probability 0 or 1. Because of the assumption that C is a pattern, we have $P\left(\mathcal{B}^i_0\right) > 0$. We conclude that $P\left(\bigcup_{n=0}^\infty \mathcal{B}^i_n\right) = 1$.

On the event $\mathcal{B}^i_n$, we have:

$$\begin{aligned} d\ \big(\ & A^{\eta_i}(N+n-1) \otimes \cdots \otimes A^{\eta_i}(n)A(n-1) \otimes \cdots \otimes A(0)u, \\ & A^{\eta_i}(N+n-1) \otimes \cdots \otimes A^{\eta_i}(n)A(n-1) \otimes \cdots \otimes A(0)v\ \big)\ = 0\,. \end{aligned}$$

This is a consequence of the fact that $A^{\eta_i}(N+n-1) \otimes \cdots \otimes A^{\eta_i}(n)$ is of rank 1 on the event $\mathcal{B}^i_n$. Using the inequality (6.12), we obtain:

$$d(A(N+n-1) \otimes \cdots \otimes A(0)u, A(N+n-1) \otimes \cdots \otimes A(0)v) \ \leqslant 8N\eta_i \leqslant \frac{1}{i}\,.$$

From the monotonicity of $d(.,.)$ (see Prop. 6.3.8), we obtain, on the event $\mathcal{B}^i_n$,

$$\lim_n\ d(A(n)\cdots A(0)u, A(n)\cdots A(0)v) \leqslant \frac{1}{i}\,.$$

We conclude that $\forall i$ fixed,

$$\begin{aligned}\bigcup_{n=0}^{\infty} \mathcal{B}_n^i \quad &\subset \quad \left\{\lim_{n\to+\infty} d(A(n)\cdots A(0)u, A(n)\cdots A(0)v) \leqslant \frac{1}{i}\right\} \\ &\Rightarrow \quad P\left\{\lim_{n\to+\infty} d(A(n)\cdots A(0)u, A(n)\cdots A(0)v) \leqslant \frac{1}{i}\right\} = 1\,.\end{aligned}$$

By letting i go to $+\infty$, we obtain that $d(A(n)\cdots A(0)u, A(n)\cdots A(0)v) \overset{n\to+\infty}{\longrightarrow} 0, \quad P-a.s.$ The assumptions of Lemma 6.10.1 are verified (there exists N such that $P\{\mathtt{D}(A(N)\cdots A(0)) < +\infty\} > 0$ as a direct consequence of the existence of a pattern of rank 1). It concludes the proof. ■

6.10.3 Proof of Theorem 6.8.1

We are going to prove that the existence of a unique stationary regime implies the existence of a pattern of rank 1 (Def. 6.7.7). Using Theorem 6.7.10, the proof will then be complete. Let us prove a lemma first.

Lemma 6.10.2. *We consider a finite model "$x(n+1) = A(n)x(n)$" with $A(n) \in \{A_1, \ldots, A_p\}$ and $A_i \in \mathbb{Q}_{max}^{k\times k}, i = 1, \ldots, p$. We suppose that there is a unique stationary regime. It implies*

$$d(A(n)A(n-1)\cdots A(0)x_0, A(n)A(n-1)\cdots A(0)y_0) \overset{n\to+\infty}{\longrightarrow} 0,\ P-a.s\,. \tag{6.13}$$

Equivalently, it implies η-coupling of the trajectories corresponding to different initial conditions.

Proof. We assume that Equation (6.13) is not verified. It implies, using Proposition 6.3.8, that there exists $x_0, y_0 \in \mathbb{R}^k$ and $c > 0$ such that

$$P(\mathcal{A}) > 0, \ \ \mathcal{A} = \{\lim_n d(A(n)A(n-1)\cdots A(0)x_0, A(n)A(n-1)\cdots A(0)y_0) > c\} > 0\,. \tag{6.14}$$

Let $S =< A_1, \ldots, A_p >$ be the projective semigroup generated by the matrices of the model (Def. 6.5.1). For $x \in \mathbb{R}^k$, we define $S(x) = \{\pi(Ax), A \in S\}$. If we assume that all the matrices of $< A_1, \ldots, A_p >$ are aperiodic, then the semigroup S is projectively finite as a direct consequence of Proposition 6.5.4. It implies that $S(x)$ is finite for all x. When it is not the case, we still have that $\pi < A_1, \ldots, A_p > \cap K$ is finite for all compact K of $\mathbb{PR}^{k\times k}$. It implies that $S(x) \cap K$ is finite for all compact K of $\mathbb{PR}^k$. We conclude that $S(x)$ has no accumulation point and verifies $\overline{S(x)} = S(x)$, where $\overline{S(x)}$ is the closure of $S(x)$ in $\mathbb{PR}^k$.

We want to apply Theorem 6.6.8. It is required that the probability space be a Polish space. In order to fulfill this, we consider the canonical probability space consisting of one-sided infinite sequences of matrices $\{A_1, \ldots A_p\}$, i.e.

$$\Omega = \{(A_{u_0}, A_{u_1}, \ldots, A_{u_n}, \ldots),\ u_i \in \{1, \ldots, p\}\}\,.$$

When the set $S(x)$ is finite, it is immediate that the sequence $\{\pi(x(n,x))\}$ is tight in $S(x)$. Let us prove it is still true in the general case. We recall that we made the assumption (6.6), which implies

$$\lim_n P\left\{\mathtt{D}(\, A(n)\cdots A(0)\,) < +\infty\right\} = 1\,.$$

It implies that for all $\eta > 0$, there exists $N \in \mathbb{N}$ and K, a compact set of $\mathbb{PR}^{k\times k}$ such that

$$\forall n \geqslant N,\ P\left\{\pi(A(N)\cdots A(0)) \in K\right\} > 1-\eta\,.$$

There exists a compact K' (which depends on x) of $\mathbb{PR}^k$ such that

$$\{\pi(A(N)\cdots A(0)) \in K\} \Rightarrow \{\pi(A(N)\cdots A(0)x) \in K'\}\,.$$

We conclude that the sequence $\{\pi(x(n,x))\}$ is tight in $\mathbb{PR}^k$. It implies that it is tight in $\overline{S(x)} = S(x)$. We can view $\pi(x(n,x))$ as a SRS defined on $S(x)$ only. Applying Theorem 6.6.8, we obtain that, for all x, there exists a stationary distribution Q_x defined on $\Omega \times S(x)$.

Let us consider the initial conditions x_0 and y_0 as defined in (6.14). It is a-priori possible to have $S(x_0)\cap S(y_0) \neq \emptyset$. As a consequence, one cannot rule out that $Q_{x_0} = Q_{y_0}$. We are going to prove that there exists $\alpha \in \mathbb{R}$ such that $S(x_0)\cap S(\alpha x_0 \oplus y_0) = \emptyset$. It will provide two stationary distributions $Q_{x_0} \neq Q_{\alpha x_0 \oplus y_0}$, which contradicts the uniqueness of the stationary regime.

We work on the event $\mathcal{A}$, see (6.14). We have $d(x(n,x_0),x(n,y_0)) > c$ for all n. Let $x, y \in \mathbb{R}^k$ be two different points. Then there exists an open interval $]\underline{\lambda},\overline{\lambda}[$ such that

$$\overline{\lambda} - \underline{\lambda} = d(x,y),\ \{\overline{\lambda}x \oplus y, \underline{\lambda}x \oplus y\} = \{x,y\}\,,$$

$$\forall \lambda \neq \lambda' \in [\underline{\lambda},\overline{\lambda}],\ \lambda x \oplus y \neq \lambda' x \oplus y\,.$$

The proof is straightforward, for more insights, see Chapter 3, §3.4.2. We consider the (random) intervals $]\underline{\lambda}(n),\overline{\lambda}(n)[$ defined as above for the couples of points $\{x(n,x_0),x(n,y_0)\}$. For any $A \in \mathbb{R}_{max}^{k\times k}, x, y \in \mathbb{R}_{max}^k$ and $\lambda \in \mathbb{R}$, we have $A(\lambda x \oplus y) = \lambda Ax \oplus Ay$. As a consequence, the sequence $]\underline{\lambda}(n),\overline{\lambda}(n)[$ is decreasing. Let $\underline{\lambda}$ and $\overline{\lambda}$ be the limits of $\underline{\lambda}(n)$ and $\overline{\lambda}(n)$. On the event $\mathcal{A}$, we have $\overline{\lambda} - \underline{\lambda} > c$ (see (6.14)).

We define the sets

$$\Lambda(n) = \{\lambda \mid\ \pi(\lambda x(n,x_0) \oplus x(n,y_0)) \in S(x_0)\},\ \Lambda = \bigcup_{n\in\mathbb{N}} \Lambda(n)\,.$$

Let $x, y, z \in \mathbb{PR}^k$ be three different points. It is immediate to prove that there exists a unique $\lambda \in \mathbb{R}$ such that $z = \lambda x \oplus y$. As a consequence, the sets $\Lambda(n)$ are countable and Λ is countable. It implies that the set $]\underline{\lambda},\overline{\lambda}[\backslash\Lambda$ is non-empty on $\mathcal{A}$. For all $\lambda \in]\underline{\lambda},\overline{\lambda}[\backslash\Lambda$, we have, by definition of Λ, that $S(\lambda x_0 \oplus y_0)\cap S(x_0) = \emptyset$. The conclusion follows. □

Remark The proof does not work when matrices $A_1, \ldots, A_p$ belong to $\mathbb{R}_{max}^{k\times k}$. In this case, it is possible to have $\overline{S(x)} \neq S(x)$. In the model detailed in Example 6.8.2, all the sets $S(x)$ are dense in the interval $[-1,1]$ (as a classical consequence of the assumption $\eta/\eta' \notin \mathbb{Q}$). It implies that $\overline{S(x)} = [-1,1], \forall x$. The stationary distributions Q_x are all defined on the same set, $\Omega \times [-1,1]$, which prevents the previous proof from working.

We want to prove the existence of a rank 1 pattern of $\{A(n)\}$ (Def. 6.7.7). There exists a r.v. N such that $A(N)\cdots A(0)_{ij} > \varepsilon, \forall i,j$ (consequence of Equation (6.6)).
It follows from the ergodic Lemma 10.2.1, that the set

$$\mathcal{I} = \{n \mid n \geqslant N,\ A(n)\cdots A(n-N) = A(N)\cdots A(0)\} \tag{6.15}$$

is infinite, P-a.s. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be the strictly increasing function such that $\mathcal{I} = \{\sigma(0), \sigma(1), \ldots\}$. We define the subsequence $\{B(n) = A(\sigma(n))A(\sigma(n)-1)\cdots A(0),\ n \in \mathbb{N}\}$. The matrices $B(n)$ can be written under the form $B(n) = A(N)\cdots A(0)\tilde{B}(n)A(N)\cdots A(0)$ for $n \geqslant 3$. We have

$$\begin{aligned} \max_{ij} B(n)_{ij} &\leqslant \max_{ij} A(N)\cdots A(0)_{ij} \otimes \max_{ij} \tilde{B}(n)_{ij} \otimes \max_{ij} A(N)\cdots A(0)_{ij} \\ &\leqslant \max_{ij} A(N)\cdots A(0)_{ij} \otimes \tilde{B}(n)_{uv} \otimes \max_{ij} A(N)\cdots A(0)_{ij}\ , \end{aligned} \tag{6.16}$$

for some indices u, v belonging to the argmax in $\max_{ij} \tilde{B}(n)_{ij}$. We also have

$$\begin{aligned} \forall i,j,\ B(n)_{ij} &\geqslant A(N)\cdots A(0)_{iu} \otimes \tilde{B}(n)_{uv} \otimes A(N)\cdots A(0)_{vj} \\ \min_{ij} B(n)_{ij} &\geqslant \min_{ij} A(N)\cdots A(0)_{ij} \otimes \tilde{B}(n)_{uv} \otimes \min_{ij} A(N)\cdots A(0)_{ij}\ . \end{aligned} \tag{6.17}$$

We consider the Euclidean space $(\mathbb{PR}^{k\times k}, |.|_{\mathcal{P}})$ where $|.|_{\mathcal{P}}$ is the norm introduced in Definition 6.3.7. It follows from (6.16) and (6.17) that

$$\begin{aligned} |B(n)|_{\mathcal{P}} = \max_{ij} B(n)_{ij} - \min_{ij} B(n)_{ij} &\leqslant 2 \times (\max_{ij} A(N)\cdots A(0)_{ij} - \min_{ij} A(N)\cdots A(0)_{ij}) \\ &= 2 \times |A(N)\cdots A(0)|_{\mathcal{P}}\ . \end{aligned}$$

It implies that the sequence $\{\pi(B(n))\}$ belongs to a compact of $(\mathbb{PR}^{k\times k}, |.|_{\mathcal{P}})$. Hence there exists a strictly increasing function $\sigma : \mathbb{N} \to \mathbb{N}$ such that $\pi(B(\sigma(n)))$ is converging. Let A_∞ be a representative (in $\mathbb{R}^{k\times k}$) of the limit. By continuity of the projective distance, we have that $\forall u, v \in \mathbb{R}^k,\ d(A_\infty u, A_\infty v) = 0$. Therefore A_∞ is a rank 1 matrix.

As the products $\{\pi(A(n)\cdots A(0))\}$ can only take a finite number of values in compact sets (Proposition 6.5.4), it implies that the limit matrix A_∞ is attained in finite time. More precisely, there exists N such that

$$\forall n \geqslant N,\ \pi\left(B(\sigma(n))\right) = \pi(A_\infty)\ .$$

The matrix $B(\sigma(N))$ is a rank 1 pattern for $\{A(n)\}$. It concludes the proof. ∎

6.10.4 Proof of Theorem 6.8.4

We first prove the necessary part of the Theorem, i.e. $\eta-$coupling with a unique stationary regime implies the existence of an asymptotic pattern.
Let $Z \circ \theta^n$ be the unique stationary regime. We have for all $x_0, y_0 \in \mathbb{R}^k$,

$$d(x(n,x_0), Z \circ \theta^n) \to 0,\ d(x(n,y_0), Z \circ \theta^n) \to 0 \ \Rightarrow\ d(x(n,x_0), x(n,y_0)) \to 0\ .$$

We have assumed that $\exists N$ such that $P\{\mathtt{D}(A(N-1)\cdots A(0)) < +\infty\} > 0$, see Equation (6.6), Section §6.8. Let $K \in \mathbb{R}$ be such that $P\{\mathtt{D}(A(-1)\cdots A(-N)) < K\} > 0$. It implies that there exists K' such that $P\{|A(-1)\cdots A(-N)|_{\mathcal{P}} < K'\} > 0$. Let us denote

$$\mathcal{E}_0 = \{\omega \mid\ \ |A(-1)\cdots A(-N)|_{\mathcal{P}} < K'\}\ .$$

It follows from the stationary-ergodic assumption, that there exists a minimal $n_1 > 1$ such that

$$P\{\mathcal{E}_1\} > 0,\ \mathcal{E}_1 = \mathcal{E}_0 \cap \{|A(-n_1)\cdots A(-n_1 - N + 1)|_{\mathcal{P}} < K'\}\ .$$

We define in the same way an increasing sequence $n_p > \cdots > n_2 > n_1$ and a decreasing sequence of events $\mathcal{E}_p \subset \cdots \subset \mathcal{E}_2 \subset \mathcal{E}_1$ verifying

$$P\{\mathcal{E}_p\} > 0,\ \mathcal{E}_p = \mathcal{E}_{p-1} \cap \{|A(-n_p)\cdots A(-n_p - N + 1)|_{\mathcal{P}} < K'\}\ .$$

On the event $\mathcal{E}_p, p \geqslant 1$, we have

$$|A(-1)\cdots A(-n_p - N + 1)|_{\mathcal{P}} < 2 \times K'\ .$$

The proof is exactly similar to the one proposed in the proof of Theorem 6.8.1 (§6.10.3, Equation (6.15) and after). Let $\mathcal{B}(E, K')$ denote the open ball of $(\mathbb{P}\mathbb{R}^{k\times k}, |.|_{\mathcal{P}})$ of center $\pi(E), E_{ij} = e, \forall i,j$ and of radius K'. For all p, we choose a deterministic matrix B_p belonging to $\mathcal{B}(E,K')$ and verifying

$$P\left\{\{\mathcal{E}_p\} \cap \{\pi A(-1)\cdots A(-n_p - N + 1) \in \mathcal{B}(B_p, \frac{1}{p})\}\right\} > 0\ . \tag{6.18}$$

As the matrices $\{B_p, p \in \mathbb{N}\}$ belong to a compact, there exists a subsequence $\{B_{\sigma(p)}\}$ which converges to a limit B_∞. We have (see the proof of Lemma 6.10.1) that $\mathtt{D}(A(-1)\cdots A(-n)) \to 0,\ P-a.s.$. We conclude that B_∞ is a rank 1 matrix.
We fix $\eta > 0$. Let C be such that $\forall p > C$, we have $|B_\infty - B_{\sigma(p)}|_{\mathcal{P}} \leqslant \eta/2$. For $p > \max(C, 2/\eta)$, we have $\mathcal{B}(B_{\sigma(p)}, 1/p) \subset \mathcal{B}(B_\infty, \eta)$. It implies

$$P\{\pi A(-1)\cdots A(-n_{\sigma(p)} - N + 1) \in \mathcal{B}(B_\infty, \eta)\} \geqslant$$
$$P\left\{\{\mathcal{E}_{\sigma(p)}\} \cap \{\pi A(-1)\cdots A(-n_{\sigma(p)} - N + 1) \in \mathcal{B}(B_{\sigma(p)}, \frac{1}{p})\}\right\} > 0\ .$$

It means precisely that B_∞ is an asymptotic pattern of $\{A(n)\}$, see Definition 6.7.13.

Let us prove the sufficient part of the theorem. We assume that there exists a deterministic matrix $\tilde{A}$ which is a rank 1 asymptotic pattern of $\{A(n)\}$. We want to prove the η-coupling convergence of $\pi(x(n))$ to a unique stationary regime.
We fix $\eta > 0$. Let N_η be such that

$$P\left\{\pi A(N_\eta - 1)\cdots A(0) \in \mathcal{B}(\tilde{A}, \eta)\right\} > 0\,. \tag{6.19}$$

Using the ergodic Lemma 10.2.1, we have

$$P\{\exists i \geqslant 0 \mid \ \pi A(N_\eta - 1 + i)\cdots A(i) \in \mathcal{B}(\tilde{A}, \eta)\} = 1\,.$$

Let u be the unique eigenvector of the rank 1 matrix $\tilde{A}$ and $\mathcal{B}(u, \eta)$ the ball of center $\pi(u)$ and radius η in $\mathbb{PR}^k$. We have that for all $x_0 \in \mathbb{R}^k$,

$$\{\pi(x(n, x_0)) \in \mathcal{B}(u, \eta)\} \subset \{\pi A(n-1)\cdots A(n - N_\eta) \in \mathcal{B}(\tilde{A}, \eta)\}\,.$$

In particular, it implies that $\forall x_0, y_0 \in \mathbb{R}^k$ and n large enough,

$$\{d(x(n, x_0), x(n, y_0)) < \eta\} \subset \bigcup_{N_\eta \leqslant p \leqslant n} \{\pi A(p-1)\cdots A(p - N_\eta) \in \mathcal{B}(\tilde{A}, \eta)\}\,.$$

We deduce that $P\{d(x(n, x_0), x(n, y_0)) < \eta\} \to 1$. We conclude by using Lemma 6.10.1 (the existence of n such that $P\{\mathtt{D}(A(n)\cdots A(0) < +\infty\} > 0$ comes from Equation (6.19)).

■

6.10.5 Proof of Theorem 6.8.5

We want to prove that the conditions given in Theorem 6.8.5 are necessary. We suppose that our model couples in finite time with a unique stationary regime, uniformly over initial conditions in $\mathbb{R}^k$. Let us prove a lemma first.

Lemma 6.10.3. *If there is a unique stationary regime for $\pi(x(n))$, coupling in finite time uniformly over initial conditions in $\mathbb{R}^k$ implies strong coupling in finite time uniformly over initial conditions in $\mathbb{R}^k$.*

Proof. Let $\{Z \circ \theta^n\}$ be the unique stationary regime with which the SRS $\pi(x(n))$ couples. We consider the event:

$$\mathcal{Y}_n = \left\{\omega \mid \pi(x(n, x_0)) \circ \theta^{-n}\omega = Z\omega,\ \ \forall x_0 \in \mathbb{R}^k\right\}\,.$$

The assumption of coupling in finite time, uniformly over $\mathbb{R}^k$, may be written :

$$P(\mathcal{Y}_n) \overset{n\to+\infty}{\longrightarrow} 1\,.$$

Here we implicitly use the assumption that the projective image of $A(-1)\cdots A(-n)$ is asymptotically bounded (see Equation (6.6)). Let us consider $\omega \in \mathcal{Y}_n$ and p an integer > 0, we have:

$$\begin{aligned} \pi(x(n+p,x_0)) \circ \theta^{-n-p}\omega &= \pi\left(x(n, x(p,x_0)\circ\theta^{-p}) \right) \circ \theta^{-n}\omega \qquad (6.20) \\ &= Z\omega \quad (\text{as } \omega \in \mathcal{Y}_n)\,. \qquad (6.21) \end{aligned}$$

The passage from (6.20) to (6.21) uses the fact that coupling occurs uniformly over initial conditions. We have:

$$\mathcal{Y}_n = \left\{ \omega \mid \pi(x(n+p,x_0)) \circ \theta^{-(n+p)}\omega = Z\omega,\ \ \forall p \geqslant 0,\ \ \forall x_0 \in \mathbb{R}^k \right\}\,,$$

and

$$P(\mathcal{Y}_n) \overset{n\to+\infty}{\longrightarrow} 1\,.$$

This is exactly the definition of strong coupling (Def. 6.6.4). □

We can now use the converse Theorem 6.6.7. There exists a stationary sequence of events $\{\mathcal{A}\circ\theta^n\}$ which is renovating for the SRS $\{\pi(\,x(n,x_0)\,)\}$, $\forall x_0 \in \mathbb{R}^k$, and verifies $P(\mathcal{A}) > 0$. Let m be the common length and Φ the common function of these renovating events. We have, on $\mathcal{A}$:

$$\pi(x(m)) = \Phi\left(A(m-1),\ldots,A(0)\right),\ \ \forall x_0 \in \mathbb{R}^k\,.$$

But we also have:

$$x(m) = A(m-1)\otimes\cdots\otimes A(0)\otimes x_0,\ \ \forall x_0 \in \mathbb{R}^k\,.$$

We conclude that, on $\mathcal{A}$, $\pi\left(A(m-1,\omega)\otimes\cdots\otimes A(0,\omega)\otimes x_0\right)$ is independent of x_0. It implies that $C = A(m-1,\omega)\otimes\cdots\otimes A(0,\omega)$ is a matrix of rank 1. ■

We have not represented in Figure 6.4 all the implications which are true, but only the ones which are used in the proofs. The exclamation mark "!" stands for a trivial proof, the question mark "Γ" for an open question. For finite models with matrices in $\mathbb{Q}_{max}^{k\times k}$, the diagram is completely commutative. We have the necessary and sufficient conditions of uniqueness of the stationary regime. For general models, we have two commutative sub-diagrams. They correspond to the necessary and sufficient conditions of stability with $\eta-$coupling and coupling respectively.

Unsolved problems Here are two questions which we have not been able to solve. They would enable to complete the stability picture as shown in Figure 6.4.

1. Is Lemma 6.10.2 still true if we consider a finite model with matrices in $\mathbb{R}_{max}^{k\times k}$ Γ Is it still true for a general model Γ

2. For a finite and i.i.d. model with matrices in $\mathbb{R}_{max}^{k\times k}$, do we have that the existence of an asymptotic pattern of rank 1 implies the existence of a pattern of rank 1 Γ The counter-example proposed in Example 6.8.2, was for a finite and stationary-ergodic model with matrices in $\mathbb{R}_{max}^{k\times k}$.

If the answer to both questions is positive, then we can extend Theorem 6.8.1 to finite and i.i.d. models with matrices in $\mathbb{R}_{max}^{k\times k}$.

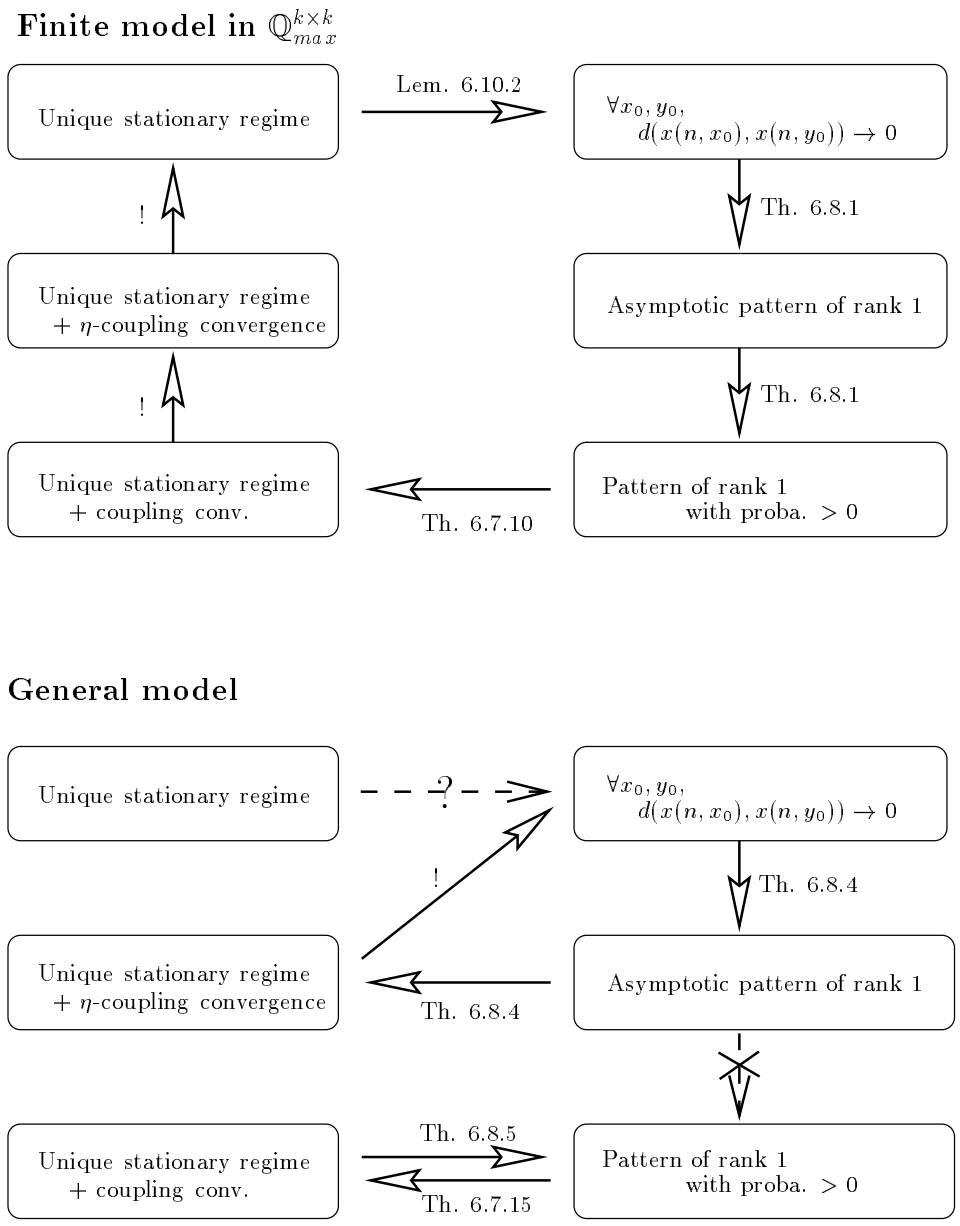


Figure 6.4: Summary of the results.

Chapitre 7

Illustration of Multiple Stationary Regimes

Illustration des Régimes Stationnaires Multiples

Ce chapitre est une illustration des résultats du chapitre 6 à l'aide de l'outil graphique du chapitre 3.

The purpose of this Chapter is to study the uniqueness or multiplicity of stationary regimes for stochastic (max,+) linear systems. We will see that stability is by far the most common situation. We obtain multiple stationary regimes when the system is not "stochastic enough" in some sense. The main cases of multiplicity will be illustrated using the graphical representation of Chapter 3.

7.1 Introduction

We consider a stochastic (max,+) linear system

$$\left\{ \begin{array}{rcl} x(n+1) & = & A(n)x(n), \quad n \in \mathbb{N} \\ x(0) & = & x_0 \, , \end{array} \right. \tag{7.1}$$

The matrices $A(n) \in \mathbb{R}_{max}^{k \times k}$ verify the minimal assumptions of §6.9.1. We deduce from Theorems 6.8.1 and 6.8.5, that a good way to show the stability of system (7.1) is to extract some deterministic matrices from the support of $A(0)$ and to build a product of these matrices which is scs1-cyc1. In most cases, an extracted model with two matrices is enough to conclude. In the following, we are going to illustrate the phenomena of uniqueness or multiplicity of stationary regimes with models of two matrices.

From now on, when no other specific assumption is made, it is assumed that $A(n) = A$ with probability $p > 0$ and $A(n) = B$ with probability $1 - p > 0$, A and B being irreducible deterministic matrices. It is also assumed that the sequence $\{A(n)\}$ is i.i.d.

7.2 Multiple Stationary Regimes

It is clear that there are several stationary regimes when the sets of eigenvectors of matrices A and B contain more than one point. Let us propose two examples.

Example 7.2.1. We consider in Figure 7.1.I., the matrices :

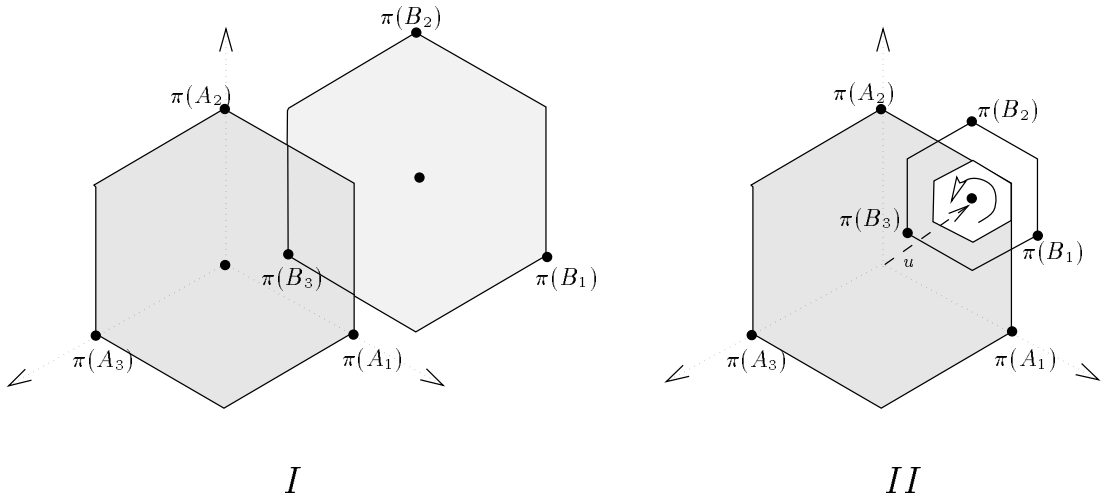
$$A = \begin{pmatrix} e & -1 & -1 \\ -1 & e & -1 \\ -1 & -1 & e \end{pmatrix}, \quad B = PAP^{-1} \text{ with } P = \begin{pmatrix} 1.4 & \varepsilon & \varepsilon \\ \varepsilon & 1.3 & \varepsilon \\ \varepsilon & \varepsilon & e \end{pmatrix} .$$

The intersection Π of the sets of eigenvectors of A and B is not empty. It implies a continuum of stationary regimes. Indeed if we consider an initial condition $x_0 \in \Pi$, we have $\pi(x(n, x_0)) = \pi(x_0), \forall n$.

Example 7.2.2. We consider in Figure 7.1.II., a system with the same matrix A as above and

$$B = \begin{pmatrix} 0.7 & \varepsilon & \varepsilon \\ \varepsilon & 0.8 & \varepsilon \\ \varepsilon & \varepsilon & e \end{pmatrix} \begin{pmatrix} -0.5 & -0.5 & e \\ e & -0.5 & -0.5 \\ -0.5 & e & -0.5 \end{pmatrix} \begin{pmatrix} -0.7 & \varepsilon & \varepsilon \\ \varepsilon & -0.8 & \varepsilon \\ \varepsilon & \varepsilon & e \end{pmatrix} .$$

Matrix B has been written under a form which emphasizes that B is obtained as a translation (see §3.4.3) of a canonical scs1-cyc3 matrix. Let $\mathcal{E}_A$ be the set of eigenvectors of A and $\mathcal{E}_B$



I *II*

Figure 7.1: I : Example 7.2.1. II : Example 7.2.2

the set of periodic regimes (of period 3) of B. Let $u = (0.7, 0.8, e)'$ be the unique eigenvector of B. The vector u is in the interior of the closed polyhedral set $\mathcal{E}_A \cap \mathcal{E}_B$. There is a ball $\mathcal{B}(\pi(u), \beta)$ (for the projective distance) centered in $\pi(u)$ and of radius $\beta > 0$ which is contained in $\pi(\mathcal{E}_A \cap \mathcal{E}_B)$. We consider a new basis of center u. For an initial condition $x = (x_1, x_2, x_3)'$ such that $\pi(x) \in \mathcal{B}(\pi(e), \beta)$ (in the new basis), the state space of the Markov chain $\pi(x(n))$ is

$$\left\{\pi\left(\sigma^n(x_1, x_2, x_3)'\right),\ n = 1, 2, 3\right\} \text{ with } \sigma(x_1, x_2, x_3)' = (x_2, x_3, x_1)'.$$

Let us denote $\alpha_i = \pi\left(\sigma^i(x_1, x_2, x_3)'\right)$. The probability transitions of the Markov chain $\pi(x(n))$ are $p(\alpha_i, \alpha_{i+1}) = 1 - p,\ p(\alpha_i, \alpha_i) = p$.

Under the light of the previous examples, a natural conjecture would be the following one

Let A and B be two irreducible matrices. There is a finite product of A and B which is scs1-cyc1 if and only if the sets of eigenelements (eigenvectors + periodic regimes) of A and B have an intersection which is empty or restricted to one point.

However, this result is false as illustrated by Example 7.2.3.

Example 7.2.3. In Figure 7.2, we consider :

$$A = \begin{pmatrix} e & . & . \\ . & e & . \\ . & . & -2 \end{pmatrix}, (.) = -1, \quad B = PAP^{-1} \text{ with } P = \begin{pmatrix} e & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon \\ \varepsilon & \varepsilon & 0.5 \end{pmatrix}.$$

Both matrices are scs2-cyc1. The set of eigenvectors of B is obtained from the one of A by a translation of $\pi(e, e, 0.5)'$. There is a one to one correspondence between eigenvectors of

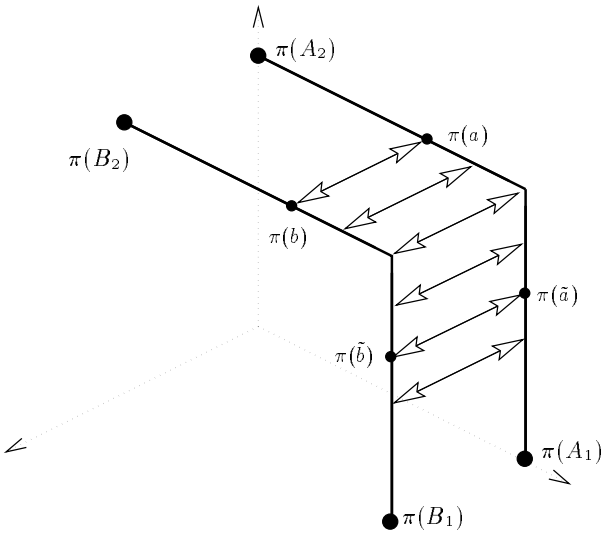


Figure 7.2: Matrices A and B are scs2-cyc1. Continuum of stationary regimes.

A and B, see the domains of attraction of scs2-cyc1 matrices as illustrated in §3.5.3, Figure 3.8. Let $a = (a_1, a_2, a_3)'$ be an eigenvector of A and $b = (a_1, a_2, a_3 + 0.5)$ the corresponding eigenvector of B. If we consider an initial condition $x_0 = a$, $\pi(x(n))$ is a Markov chain over the state space $\{\pi(a), \pi(b)\}$ with transition probabilities $P(\pi(a), \pi(a)) = p, P(\pi(a), \pi(b)) = 1 - p, P(\pi(b), \pi(b)) = 1 - p, P(\pi(b), \pi(a)) = p$.

To further illustrate the complexity of the phenomena which are involved, let us consider a slight modification of Example 7.2.3.

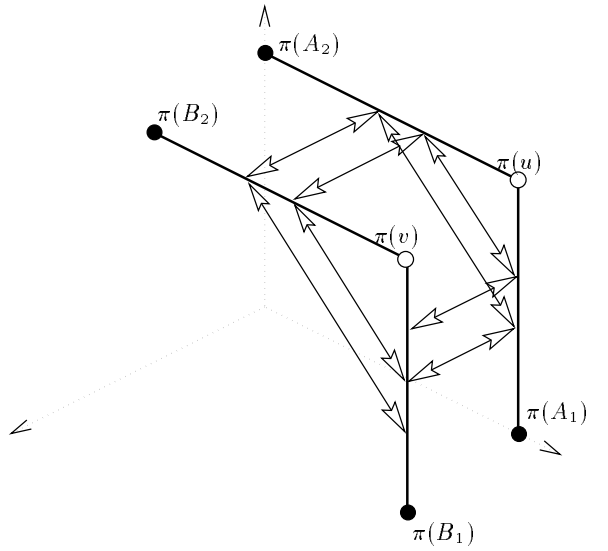


Figure 7.3: Matrices A and B are scs1-cyc2. Unique stationary regimes.

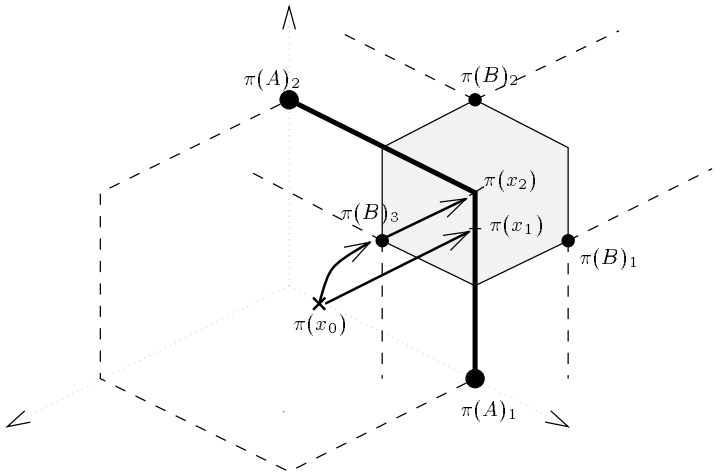


Figure 7.4: The Markov chain $\pi x(n, x_0)$ has two absorbing sets.

Example 7.2.4. In Figure 7.3, we consider :

$$A = \begin{pmatrix} . & e & . \\ e & . & . \\ . & . & -2 \end{pmatrix}, (.) = -1, \ B = PAP^{-1} \text{ with } P = \begin{pmatrix} 0.1 & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon \\ \varepsilon & \varepsilon & 0.5 \end{pmatrix}.$$

Both matrices are scs1-cyc2. The set of periodic regimes of matrix B is the one of A translated by $\pi(0.1, e, 0.5)$. If we had considered a translation of $\pi(e, e, 0.5)$, we would have obtained exactly the same Figure as in 7.2, with periodic regimes instead of eigenvectors. There would have been multiple stationary regimes. For example, the set $\{\pi(a), \pi(b), \pi(\tilde{a}), \pi(\tilde{b})\}$ would have been a possible state space for the Markov chain $\pi(x(n))$. On the other hand, with a translation of $\pi(0.1, 0, 0.5)$, we obtain a unique stationary regime as illustrated in Figure 7.3. The points $\pi(A_i)$ or $\pi(B_i), i = 1, 2$, are regenerative points for the Markov chain $\pi(x(n))$.

We have presented above some examples of different stationary regimes corresponding to different initial conditions. Here is another problem worth considering : what happens for a fixed deterministic initial condition x_0 Γ Is it possible for the Markov Chain $\pi x(n, x_0)$ to be transient Γ to have several classes of recurrence Γ The answer to both questions is positive.

Example 7.2.5. In Figure 7.4, we consider :

$$A = \begin{pmatrix} e & -2 & -2 \\ -2 & e & -2 \\ -2 & -2 & -4 \end{pmatrix}, \ B = \begin{pmatrix} e & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon \\ \varepsilon & \varepsilon & -2 \end{pmatrix} \begin{pmatrix} e & . & . \\ . & e & . \\ . & . & e \end{pmatrix} \begin{pmatrix} e & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon \\ \varepsilon & \varepsilon & 2 \end{pmatrix}, (.) = -1 .$$

The intersection between the sets of eigenvectors of matrices A and B is not empty, hence there are multiple stationary regimes. We consider the initial condition $x_0 = (0.5, e, e)'$. By direct computation, we obtain that $\pi A x_0 = \pi x_1 = \pi(2, 1.5, e)'$ and $\pi AB x_0 = \pi AB^n x_0 = \pi x_2 = \pi(2, 2, e)'$. Vectors x_1 and x_2 are common eigenvectors of A and B. We conclude that $\lim_n P(\pi x(n, x_0) = \pi x_1) = p$ and $\lim_n P(\pi x(n, x_0) = \pi x_2) = 1 - p$.

Example 7.2.6. We consider a system verifying Equation (7.1) with

$$A(n) = \begin{pmatrix} U(n) & e \\ e & U(n) \end{pmatrix} ,$$

where the random variables $U(n)$ are i.i.d and uniform over $[1, 2]$. For an initial condition $x_0 = (a, b)'$ with $a \geqslant b + 2$, we have $\pi x(n, x_0) = \pi(\inf_{k \leqslant n} U(k), 0)'$. We have a transient Markov chain.

Example 7.2.7. We can mix the previous examples. For a given initial condition, we can obtain a transient Markov chain with probability p and a recurrent Markov chain with probability $1 - p$.
More precisely, we consider a system verifying Equation (7.1). With probability p, $A(0) = A$ where A is the matrix of Example 7.2.5, and with probability $1 - p$, $A(0)$ has the following law:

$$A(0) = \begin{pmatrix} U(0) & e & 4 \\ e & U(0) & 4 \\ -4 & -4 & U(0) \end{pmatrix} ,$$

where the random variable $U(0)$ is uniform over $[0.5, 1]$. We consider the initial condition $(0.5, e, e)'$.

7.3 Quantitative Results

7.3.1 Systems with two matrices

Theorem 7.3.1. *Let matrices A and B have the following properties :*

1. *k, the size of the matrices, is prime.*
2. *A and B have a critical circuit of length k.*
3. *$\pi(u_1) \neq \pi(u_2)$ where u_1 and u_2 are the (unique) eigenvectors of A and B respectively.*

There exists a finite product of matrices A and B which is scs1-cyc1. Equivalently, the associated stochastic system has a unique stationary regime.

Before proving it, let us show that Theorem 7.3.1 fails to be true when we try to relax some of the assumptions.

Example 7.3.2. [A and B have the same eigenvector]

We denote by u the common eigenvector of A and B. Let us define

$$\alpha = \lim_l \sup_{v \in \mathbb{R}^k} d(A^l v, u) \, , \beta = \lim_l \sup_{v \in \mathbb{R}^k} d(B^l v, u) \, .$$

The balls $\mathcal{B}(\pi u, \alpha)$ and $\mathcal{B}(\pi u, \beta)$ are the sets of periodic regimes of matrices A and B respectively. We set $\gamma = \min(\alpha, \beta)$. It is easy to check that the ball $\mathcal{B}(u, \gamma)$ is a set of periodic regimes of period k for both matrices. We propose an example on Figure 7.5.

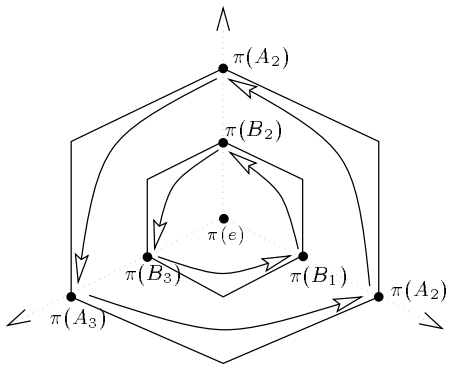


Figure 7.5: Matrices A and B are scs1-cyc3 with the same eigenvector, $\pi(e)$. Multiplicity of stationary regimes.

Example 7.3.3. [k is not prime]

We consider :

$$A = \begin{pmatrix} . & . & . & e \\ e & . & . & . \\ . & e & . & . \\ . & . & e & . \end{pmatrix}, \quad B = \begin{pmatrix} . & . & . & -1 \\ 1 & . & . & . \\ . & -1 & . & . \\ . & . & 1 & . \end{pmatrix}, (.) = -3 .$$

The eigenvectors of A and B are $(e, \ldots, e)'$ and $(-1, e, -1, e)'$ respectively. We consider the initial vector $x_0 = (a, b, a, b)'$ with $a - 2 \leqslant b \leqslant a + 2$. We verify easily that the Markov chain $\pi(x(n))$ evolves in the state space

$$\left\{ \pi(a,b,a,b)', \pi(b,a,b,a)', \pi(a \overset{-}{+} 1, b \overset{+}{-} 1, a \overset{-}{+} 1, b \overset{+}{-} 1)', \pi(b \overset{-}{+} 1, a \overset{+}{-} 1, b \overset{-}{+} 1, a \overset{+}{-} 1)' \right\} .$$

There is an infinite number of stationary regimes.

Here is another counter-example of the same kind. Suppose k is not prime. Let $1 < l < k$ be a divisor of k. We consider $(b_1, \ldots, b_l) \not\simeq (e, \ldots, e)$ and $\sum_i b_i = 0$. We define A and B in the following way: A and B have $(1, 2, \ldots, k, 1)'$ as unique critical circuit. The non-critical elements of A and B are -1. The critical elements of A and B are $(e, \ldots, e)$ and $(b_1, b_2, \ldots, b_l, b_1, b_2, \ldots, b_l, \ldots, b_1, b_2, \ldots, b_l)$ respectively.

Proof of Theorem 7.3.1

In the following, indices have to be interpreted modulo k. For example,

$$b_{k+1} = b_1 \text{ or } A_{k+3,k+2} = A_{3,2} .$$

Let λ_A and λ_B be the eigenvalues of A and B. We normalize the matrices, i.e. we set $A := A - \lambda_A$ and $B := B - \lambda_B$ (we keep the notations A and B for simplicity). Let u be the unique eigenvector

of A. We consider the matrix of change of basis P defined by $P_{ii} = u_i$ and $P_{ij} = \varepsilon, i \neq j$, see Lemma 3.4.14. We set $A := P^{-1}AP$ and $B := P^{-1}BP$. By permuting the coordinates, Lemma 3.4.13, we come down to the case where the critical circuit of A is $(1, 2, \ldots, k)$. We have :

$$A = \begin{pmatrix} . & . & . & e \\ e & . & . & . \\ . & \ddots & . & . \\ . & . & e & . \end{pmatrix} .$$

The dots (.) correspond to terms which are less or equal than e by Lemma 3.4.15. Let us assume there exists $p, q, p \neq (q+1)$ such that $A_{pq} = e$. It implies that the circuit $(q, p, p+1, \ldots, q-1)$ has a mean weight equal to e, hence is critical. Let l be the length of this circuit. We have $\text{ppcm}(l, k) = 1$ as k is prime. By Theorem 3.4.7, it implies that matrix A is scs1-cyc1 which completes the proof. In the following, we assume that $A_{pq} < e, \forall (p, q), p \neq (q+1)$.

Let us consider matrix B. If matrix B has more than one critical circuit, then it is scs1-cyc1, see above. We assume that matrix B has a unique critical circuit. We assume for the moment (and for the sake of simplicity) that is is the same as the one of A, $(1, 2, \ldots, k)$.
Let v be the unique eigenvector of B. Let $\tilde{P}$ be the matrix of change of basis associated with v, i.e. $\tilde{P}_{ii} = v_i$ and $\tilde{P}_{ij} = \varepsilon, i \neq j$. We set $\tilde{B} = \tilde{P}^{-1}B\tilde{P}$. We have $\tilde{B}_{i+1,i} = e$ and $\tilde{B}_{p,q} < e, p \neq (q+1)$, see above. We define matrix V such that $V_{ij} = v_i - v_j$. The equalities $B = \tilde{P}\tilde{B}\tilde{P}^{-1}, B^n = \tilde{P}\tilde{B}^n\tilde{P}^{-1}$ can be rewritten as ("+" is the usual sum of matrices) :

$$B = \tilde{B} + V,\ B^n = \tilde{B}^n + V, \forall n\ . \tag{7.2}$$

Matrix V is anti-symmetrical, $V_{ij} = -V_{ji}$ and $V_{ii} = e$. Let us denote $M = \max_{ij} V_{ij}$. If $M = e$, it implies that $V_{ij} = e, \forall i, j$. By Equation (7.2), we obtain $B = \tilde{B}$, which means that $e = (e, \ldots, e)'$ is the eigenvector of B. It contradicts assumption 3. of Theorem 7.3.1. We conclude that $M > e$.
We choose (i_0, j_0) such that $V_{i_0 j_0} = M$. We consider the sub-diagonal (see Definition 3.5.1) of V associated with (i_0, j_0), i.e

$$\{V_{i_0 j_0}, V_{i_0+1, j_0+1}, \ldots, V_{i_0+k-1, j_0+k-1}\}\ .$$

We set $l = k + i_0 - j_0 - 1$. The matrix $\tilde{B}^l$ has a critical sub-diagonal which is $(1, 1+l, 1+2\times l, \ldots)$. It coincides with the previously chosen sub-diagonal of V. We conclude that the maximal terms of matrix $B^l = \tilde{B}^l + V$ all are on the same sub-diagonal[1]. We denote by $\mathcal{J}$ the set of columns of B^l containing a maximal term. We sort the elements of $\mathcal{J}$ in increasing order : $\mathcal{J} = \{j_1 < j_2 < \cdots < j_{|\mathcal{J}|}\}$. We proceed in the following way.

1. For each $j_n \in \mathcal{J}$, we define the matrix $C_n = A^{k-l-j_n+1}B^l A^{j_n-1}$.

[1] When $i_0 > j_0$, it is possible to consider $l = i_0 - j_0 - 1$. The choice of $l = k + i_0 - j_0 - 1$ enables to always have $l \geqslant 0$.

2. We consider $C = C_1 C_2 \cdots C_{|\mathcal{J}|}$.

In order to illustrate and explain the choice of matrix C, we consider an example of dimension 5 with $l = 2$.

$$B^2 = \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ M & . & . & . & . \\ . & M & . & . & . \\ . & . & M & . & . \end{pmatrix}, AB^2 = \begin{pmatrix} . & . & M & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ M & . & . & . & . \\ . & M & . & . & . \end{pmatrix}, B^2A = \begin{pmatrix} . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & M \\ M & . & . & . & . \\ . & M & . & . & . \end{pmatrix}$$

The dots (.) correspond to terms which are strictly less than M. The effect of a left multiplication by A is to translate the maximal terms down. The right multiplication by A translates the maximal terms to the left. In each of the matrices $C_1 = A^3B^2, C_2 = A^2B^2A$ and $C_3 = AB^2A^2$, there is a maximal term M in place $(1,1)$ and the other maximal terms are on the diagonal. The matrix $C = C_1C_2C_3$ has the same property, i.e $C_{11} = M^{\otimes 3}$ and $C_{ij} < M^{\otimes 3}$, $i \neq j$. In fact it is easy to see on this example that we also have $C_{ii} < M^{\otimes 3}$, $\forall i \neq 1$. Hence matrix C is scs1-cyc1.

We want to prove that the same result holds for the general matrix C defined above. By construction, we have :

$$\begin{aligned} C_{11} &= (A^{k-l-j_1+1}B^lA^{j_1-1})_{11} \cdots (A^{k-l-j_{|\mathcal{J}|}+1}B^lA^{j_{|\mathcal{J}|}-1})_{11} && (7.3) \\ &= A^{k-l-j_1+1}_{1j_1+l} B^l_{j_1+l,j_1} A^{j_1-1}_{j_11} \cdots A^{k-l-j_{|\mathcal{J}|}+1}_{1j_{|\mathcal{J}|}+l} B^l_{j_{|\mathcal{J}|}+l,j_1} A^{j_{|\mathcal{J}|}-1}_{j_{|\mathcal{J}|}1} && (7.4) \\ &= e^{k-l}Me^{j_1-1} \cdots e^{k-l}Me^{j_{|\mathcal{J}|}-1} = M^{\otimes|\mathcal{J}|} && (7.5) \end{aligned}$$

To get Equation (7.3), we use that the matrices $A^{k-l-j_n+1}B^lA^{j_n-1}$ have only one maximal term on column 1 which is the term $(1,1)$. To get Equation (7.4), we use that the matrices A^{k-l-j_n+1}, B^l and A^{j_n-1} have at most one maximal term in each column.
Let us consider $(i,j) \neq (1,1)$. It is clear that $C_{ij} \leqslant C_{11}$. We want to show that $C_{ij} < C_{11}$. As the matrices $C_n, n = 1, \ldots |\mathcal{J}|$ have all their maximal terms on the diagonal, it is immediate that $C_{ij} < C_{11}$ when $i \neq j$.
Suppose there exists $i \neq 1$ such that $C_{ii} = C_{11} = M^{\otimes|\mathcal{J}|}$. It implies that (same reasons as above for Equations (7.3) (7.4))

$$\begin{aligned} C_{ii} &= (C_1)_{ii} \cdots (C_{|\mathcal{J}|})_{ii} \\ &= A^{k-l-j_1+1}_{i,j_1+i-1+l} B^l_{j_1+i-1+l,j_1+i-1} A^{j_1-1}_{j_1+i-1,i} \cdots \\ &= e^{k-l-j_1+1} B^l_{j_1+i-1+l,j_1+i-1} e^{j_1-1} \cdots && (7.6) \end{aligned}$$

Comparing Equation (7.6) and $C_{ii} = M^{\otimes|\mathcal{J}|}$, we deduce that $B^l_{j_n+i-1+l,j_n+i-1} = M$. It implies :

$$\mathcal{J} = \{j_1, \ldots, j_{|\mathcal{J}|}\} = \{j_1 + i - 1, \ldots, j_{|\mathcal{J}|} + i - 1\} .$$

In particular, there exists $p \in \{1, \ldots, |\mathcal{J}|\}$ such that $j_1 + (i-1) = j_p$, there exists q such that $j_1 + 2 \times (i-1) = j_p + (i-1) = j_q$. We prove by recurrence that $\forall p = 1, \ldots, k, j_1 = p \times (i-1) \in \mathcal{J}$.

But $\{p \times (i-1) \,[\mathrm{mod}\, k],\ p \in \mathbb{N}\}$ is a subring of $(\mathbb{Z}/k\mathbb{Z})$. As k is prime, there exists no proper subring of $(\mathbb{Z}/k\mathbb{Z})$. We conclude that :

$$\{p \times (i-1) \,[\mathrm{mod}\, k],\ p \in \mathbb{N}\} = \mathbb{Z}/k\mathbb{Z} \Rightarrow \mathcal{J} = \mathbb{Z}/k\mathbb{Z}\,.$$

Matrix B^l has a critical cycle (the one corresponding to the sub-diagonal considered above) whose terms are all equal to M. It implies that $\pi(e,\ldots,e)'$ is an eigenvector of B^l, hence of B which is in contradiction with assumption 3. of Theorem 7.3.1. We conclude that $\forall (i,j) \neq (1,1),\ C_{ij} < C_{11}$. Matrix C is scs1-cyc1.

We have now to relax the assumption that the critical circuit of B is $(1,2,\ldots,k)$ (i.e. the same as the one of A).
In general B has a critical circuit which can be described as $(\sigma(1),\sigma(2),\sigma(3),\cdots,\sigma(k))$ where σ is a permutation of $\{1,\ldots,k\}$. Let $\hat{P}$ be the permutation matrix associated with σ, see Definition 3.4.12. Let us consider $\hat{B} = \hat{P}^{-1}B\hat{P}$. The critical circuit of $\hat{B}$ is $(1,\ldots,k)$. We associate with matrix $\hat{B}$ an integer $\hat{l}$ and a set $\hat{\mathcal{J}}$ in the same way as above. We define :

$$C_n = A^{k-\sigma(\hat{\jmath}_n+\hat{l})+1} B^{\hat{l}} A^{\sigma(\hat{\jmath}_n)-1},\ C = \prod_n C_n\,.$$

We prove that C has a unique maximal element on the diagonal, C_{11}, and the conclusion follows.

■

Remark 7.3.4. Let us propose a mathematical trivia. Consider k reals, $\{b_1,b_2,\ldots,b_k\}$ on a ring (i.e. we identify $b_{k+1}=b_1, b_{k+2}=b_2,\ldots$). We suppose that the b_i are not all equal.

Does there exist n, $1 \leqslant n \leqslant k$, such that :
Among the k partial sums of length n, $S(i) = b_i + b_{i+1} + \ldots + b_{i+n-1}$, $i = 0,\ldots,k-1$, there is one and only one maximal sum.

If the answer was positive, it would provide a very simple and elegant proof of Theorem 7.3.1. Let the critical terms of B be the reals b_i and consider the matrix B^l where l is a solution to the trivia. Matrix B^l has a unique maximal term and it is easy to obtain a matrix of the form $A^p B^l A^q$ which is scs1-cyc1.

However the answer is always negative except for $k=3$ and $k=5$! We leave the proof to the reader. Here are some hints. For $k=4$, consider $\{1,-1,1,-1\}$ and for $k=7$ consider $\{1,0,-1,1,-1,0,0\}$.

It should be possible to prove other results similar to Theorem 7.3.1 for other spectral behaviours of the matrices A and B. We feel however that entering into too much details would be of limited interest for our purpose which is to get a global understanding of the uniqueness and multiplicity of stationary regimes.

7.3.2 Other systems

We consider a different type of system. Each coordinate of the matrix $A(n)$ can take two different values. Theorem 7.3.5 illustrates the introductory remark that most stochastic systems have a unique stationary regime.

Theorem 7.3.5. *We consider a set of reals $\{(a_{ij}, b_{ij}), i, j = 1, \ldots k\}$ such that $a_{ij} < b_{ij}$, $\forall i, j$. We consider a stochastic (max,+) linear system $x(n+1) = A(n)x(n)$, where the sequence $\{A(n), n \in \mathbb{N}\}$ is i.i.d. We assume also that the coordinates $A_{ij}(0)$ are independent random variables and verify*

$$P\{A_{ij}(0) = a_{ij}\} = p_{ij} > 0,\ P\{A_{ij}(0) = b_{ij}\} = 1 - p_{ij} > 0\ .$$

This system has a unique stationary regime.

Proof. We provide only a sketch of the proof. We consider the deterministic matrix A defined by $A_{ij} = a_{ij}$.
Let us assume that there exists $p \in \{1, \ldots, k\}$ such that (p, p) is a critical circuit of A. We define the matrix C by $C_{pp} = b_{pp}$ and $C_{ij} = A_{ij} = a_{ij}, \forall (i,j) \neq (p,p)$. As $b_{pp} > a_{pp}$, we deduce that (p,p) is the unique critical circuit. Hence matrix C is scs1-cyc1. The uniqueness of the stationary regime follows.
Now we assume that there is no critical circuit of length 1. By a permutation of the coordinates, we come down to the case where $(1, 2, \ldots, l), 1 < l \leqslant k$, is a (non-necessarily unique) critical circuit of A. We define two matrices C and D in the following way :

$$C_{12} = b_{12}, C_{ij} = A_{ij} = a_{ij}, \forall (i,j) \neq (1,2),\ \ D_{l1} = b_{l1}, D_{ij} = A_{ij} = a_{ij}, \forall (i,j) \neq (l,1)\ .$$

We have two matrices with a common critical circuit, $(1, 2, \ldots, l)$. Furthermore, we have that the term C_{12} (resp. D_{l1}) belongs to any critical circuit of C (resp. D). It implies that the matrices C and D are scs1. The terms of the critical circuit $(1, 2, \ldots, l)$ are different for matrices C and D. It implies that their unique eigenvector is different. By adapting slightly the proof of Theorem 7.3.1, we obtain that there exists a finite product of matrices C and D which is scs1-cyc1. □

Chapitre 8

Application to Stochastic Event Graphs

Application aux Graphes d'Événements Stochastiques

Dans ce chapitre, nous montrons comment adapter les résultats du chapitre 6 à l'étude de la stabilité des Graphes d'Evénements stochastiques.

We show how to apply the results on stochastic (max,+) linear systems to Stochastic Event Graphs (SEG). Let $x(n+1) = A(n) \otimes x(n)$ be a (max,+) linear system representing the evolution of the SEG, see Chapter 1. We assume that the (max,+) system has a unique stationary regime. Then, under some weak additional assumptions, we prove that the SEG converges to a unique stationary regime, independently of the initial (reachable) marking.

8.1 Event Graph

For the basic definitions relative to an Event Graph, the reader is referred to Section §1.3 or §2.7. We consider a closed Event Graph (E, M). The set E corresponds to the underlying graph (the set of places, transitions and arcs). This graph is assumed to be strongly connected. Let T and P be the number of transitions and places respectively. The vector $M \in \mathbb{N}^P$ is the marking, i.e. the number of tokens in each place.
Let us recall some basic properties of Event Graphs. For more details, the reader is referred to [112] [30].

Definition 8.1.1 (Incidence matrix). *The incidence matrix G of the Event Graph is a matrix of dimension $T \times P$ defined as :*

- $G_{ij} = 1$ *if there is an arc from transition t_i to place p_j.*
- $G_{ij} = -1$ *if there is an arc from place p_j to transition t_i.*
- $G_{ij} = 0$ *otherwise.*

Definition 8.1.2 (Reachability). *A marking $\tilde{M}$ is reachable from a marking M if there is a sequence of transitions $t_0, \ldots, t_n$ and a sequence of markings $M_1, \ldots, M_n$ such that :*

$$M \xrightarrow{t_0} M_1 \xrightarrow{t_1} M_2 \cdots M_n \xrightarrow{t_n} \tilde{M} ,$$

More precisely, transition t_i is enabled in marking M_i and the firing of t_i transforms marking M_i into marking M_{i+1}.

Definition 8.1.3 (Synchronic distance). *We denote by d_{ij} the synchronic distance between two transitions t_i and t_j. It is defined as $d_{ij} = e_{ij} + e_{ji}$, where e_{ij} is the minimal number of tokens in a directed path from transition t_i to transition t_j.*

Proposition 8.1.4. *In a strongly connected Event Graph, we have the following properties*

1. *The number of tokens in a circuit is an invariant.*
2. *Marking $\tilde{M}$ is reachable from a marking M if and only if (matrix-vector product is in the usual algebra)*

$$\exists x \in \mathbb{N}^T \ : G' \times x = \tilde{M} - M , \tag{8.1}$$

where G' is the transpose of G. The vector x is called the firing count vector. The coordinate x_i corresponds to the number of firings of transition t_i in the transformation from M to $\tilde{M}$.

3. *Matrix G is of rank $T-1$. The solutions of $G'x = 0$ are $x = (n, \ldots, n)'$. The interpretation is that we go from a marking to itself if and only if each transition has fired the same number of times. As a consequence, if $\tilde{M}$ is reachable from M then M is reachable from $\tilde{M}$.*

We are going to need the following lemma.

Lemma 8.1.5. *Let us denote by K the maximal number of tokens in a circuit of (E, M). Let $\tilde{M}$ be a marking reachable from M. Let z be a minimal solution of $G' \times x = \tilde{M} - M$, $x \in \mathbb{N}^T$. We have* $\max_i z_i \leqslant K$.

Proof. We recall that $x = (n, \ldots, n)'$ verifies $G'x = 0$, Proposition 8.1.4.3. Let us assume that the minimal firing vector z is such that $\min_i z_i > 0$. Then we obtain $G' \times [z - (\min_i z_i, \ldots, \min_i z_i)] = G' \times z$ which contradicts the minimality of z. Let i_0 be such that $z_{i_0} = 0$. From the interpretation of the synchronic distance given in Definition 8.1.3, we deduce that

$$z_i = z_i - z_{i_0} \leqslant d_{i_0,i}, \ \forall i = 1, \ldots, T \, .$$

But we have $d_{ij} \leqslant K$, $\forall i, j$, as a consequence of Proposition 8.1.4.1. We conclude that $\max_i z_i \leqslant K$.

□

8.2 Stochastic Event Graph

We consider a closed FIFO Stochastic Event Graph $S = (E, M, \Sigma, Y)$. The set (E, M) is a strongly connected Event Graph, see §8.1. The sequence of firing times of transitions and holding times of places is $\Sigma = \{(a_1(n), \ldots, a_T(n)),\ n \in \mathbb{N}, (\alpha_1, \ldots, \alpha_P)\}$. The holding times are assumed to be constant in order for the FIFO assumption to be fulfilled, see [8], p. 71 or Chapter 1. The vector $Y \in \mathbb{R}^T$ is the vector of initial condition, i.e. the remaining firing time at instant zero for each transition. We set $Y_i = 0$ if transition t_i is not enabled at instant 0.

There are several equivalent ways of representing the SEG S by means of a stochastic (max,+) linear equation. Roughly speaking, a different representation is associated with each different reachable marking. For more details, see Chapter 2 §2.5 and 2.7.
We consider

$$L \ : \ \left\{ \begin{array}{rcl} x(n+1) & = & A(n)x(n) \\ x(0) & = & x_0 \end{array} \right. ,$$

a (max,+) linear system describing the evolution of the SEG S. In this modelling, the matrices $A(n)$ do not depend on the vector of lag times Y. It is only the initial condition x_0 which depends on Y, see §1 for more details.

We recall that the necessary and sufficient condition of stability of a (max,+) linear system is given in Theorem 6.8.5. We say that the SEG S is stable if the system L is stable (i.e. has a unique stationary regime). It implies in particular that the stationary distribution of $S = (E, M, \Sigma, Y)$ does not depend on the vector Y. For stable SEG, we will often omit to specify the value of Y.
We are now ready to prove the main Theorem.

Theorem 8.2.1. *We consider a closed FIFO Stochastic Event Graph $S = (E, M, \Sigma)$. Let us assume that S is stable. We assume that the sequence $\{(a_1(n), \ldots, a_T(n)), n \in \mathbb{N}\}$ is stationary and ergodic and that the sequences*

$$\{(a_1(n+n_1), a_2(n+n_2), \ldots, a_T(n+n_T)),\ n \in \mathbb{N}\}$$

are jointly stationary (i.e. have the same distribution) $\forall\ 0 \leqslant n_1 \leqslant K, \ldots, 0 \leqslant n_T \leqslant K$, where K is the maximal number of tokens in a circuit of (E, M).
Then all Stochastic Event Graphs $\tilde{S} = (E, \tilde{M}, \tilde{\Sigma})$ where $\tilde{\Sigma}$ has the same distribution as Σ and $\tilde{M}$ is a marking reachable from M, are stable. The unique stationary regime of $\tilde{S}$ has the same distribution as the one of S.

Remark 8.2.2. The condition "sequences $\{(a_1(n+n_1), a_2(n+n_2), \ldots, a_T(n+n_T)),\ n \in \mathbb{N}\}$ are jointly stationary" is verified in particular when the sequences $\{a_i(n),\ n \in \mathbb{N}\}$ are mutually independent. But it is a slightly weaker assumption than mutual independence.

Proof. Let us consider system $\tilde{S} = (E, \tilde{M}, \tilde{\Sigma})$. We consider the time evolution of $\tilde{S}$ given a vector of initial lag times $\tilde{Y}$. Let us assume that there exists an instant $d \geqslant 0$ such that the instantaneous marking $\tilde{M}(d)$ is equal to M. We block system $\tilde{S}$ at instant d. From Lemma 8.1.5, the difference in the number of firings of the transitions is less than K. From the assumption on the joint stationarity of the sequences $\{(a_1(n+n_1), \ldots, a_T(n+n_T))\}$, we deduce that system $\tilde{S}$ from instant d on has the same firing sequence (in distribution) than system S. Having the same marking and the same firing sequence, we deduce that system $\tilde{S}$ after time d is equivalent in distribution to system S.

The problem is that it is absolutely possible to have $\tilde{M}(d) \neq M$, $\forall d \geqslant 0$, i.e. the marking of S never appears during the evolution of $\tilde{S}$. Coming down to the marking of S is the basic idea of the forthcoming proof. However, we are going to use it in a more subtle way. It is possible to come down to marking M by considering the system $\tilde{S}$ at a *virtual instant* $(d_1, \ldots, d_T), d_i \geqslant 0$. The real d_i corresponds to the date of the clock associated with transition t_i. The reals d_i are not required to be equal which is the reason why we call $(d_1, \ldots, d_T)$ a virtual instant.

Let us detail the construction. To be coherent with previous notations, we denote

$$\tilde{\Sigma} = \{(\tilde{a}_1(n), \ldots, \tilde{a}_T(n)), n \in \mathbb{N}, (\alpha_1, \ldots, \alpha_P)\}\,.$$

Let us denote by $\tilde{x}_i(n)$ the n-th completion of a firing at transition t_i, for system $\tilde{S}$. By Proposition 8.1.4.3, the marking M is reachable from $\tilde{M}$. We denote by z the minimal solution of $G'x = M - \tilde{M}$. We define a virtual instant d by :

$$(d_1, \ldots, d_T) = (\tilde{x}_1(z_1), \ldots \tilde{x}_T(z_T))\,. \tag{8.2}$$

Let us consider system $\tilde{S}$ at instant d. A rigorous way to define system $\tilde{S}$ at a virtual instant is to consider the following modified system :
$\tilde{\tilde{S}} = (E, \tilde{M}, \tilde{\tilde{\Sigma}})$ with $\tilde{\tilde{\Sigma}} = \{(\tilde{\tilde{a}}_1(n), \ldots, \tilde{\tilde{a}}_T(n)), n \in \mathbb{N}, (\alpha_1, \ldots, \alpha_P)\}$ and

$$\tilde{\tilde{a}}_i(n) = \begin{cases} \tilde{a}_i(n) & \text{if } n \leqslant z_i \\ +\infty & \text{otherwise} \end{cases}\,.$$

As a consequence of Proposition 8.1.4.2, this system gets actually blocked after exactly z_i firings of transition t_i. The marking of the blocked system is M.
The sequence of firing times of $\tilde{S}$ after time d, equivalently the sequence of firing times which have not been used in the modified system $\tilde{\tilde{S}}$, is

$$\hat{\Sigma} = \{(\tilde{a}_1(n+z_1), \tilde{a}_2(n+z_2), \ldots, \tilde{a}_T(n+z_T)),\ n \in \mathbb{N}\}\ .$$

Because of the assumption on the joint stationarity of the sequences $\{(a_1(n+n_1), \ldots, a_T(n+n_T))\}$, we obtain that the distribution of $\hat{\Sigma}$ is the same as the one of $\tilde{\Sigma}$, hence the same as the one of Σ.

System $\tilde{S}$ after time d is equivalent to system $(E, M, \hat{\Sigma})$. The only difficulty is to determine the new initial condition for the system $(E, M, \hat{\Sigma})$. In particular this initial condition will not be compatible in the sense of [8] p. 70. However, system $(E, M, \hat{\Sigma})$ is equivalent in distribution to system (E, M, Σ) which, by hypothesis, converges to a unique stationary regime, independently of the initial condition. □

In Theorem 8.2.1, If we remove the assumption "sequences $\{(a_1(n+n_1), \ldots, a_T(n+n_T))\}$ are jointly stationary", Theorem 8.2.1 is not true anymore. Let us propose a counter-example.

Example 8.2.3. We work on a probability space (Ω, P, θ), where $\Omega = \{\omega_1, \omega_2\}$, $P = \{1/2, 1/2\}$ and θ is the stationary ergodic shift defined by : $\theta(\omega_1) = \omega_2$ and $\theta(\omega_2) = \omega_1$. We consider the Event Graph of Figure 8.1 with the sequence of firing times :

$$\{a_1(n, \omega_1)\} = \{3, 0, 3, 0, \ldots\},\ \{a_1(n, \omega_2)\} = \{0, 3, 0, 3 \ldots\}\ ,$$

$$\{a_2(n, \omega_1)\} = \{0, 3, 0, 3 \ldots\},\ \{a_2(n, \omega_2)\} = \{3, 0, 3, 0, \ldots\}\ .$$

The holding times are 1 on both places p_1 and p_2.

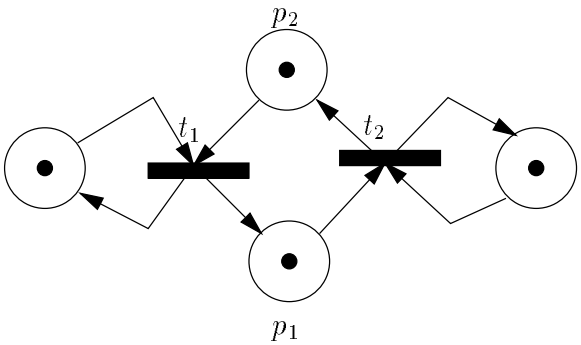


Figure 8.1: Strongly connected FIFO Event Graph.

We obtain a $\mathbb{R}_{max}$ linear representation for this model which is

$$x(n) = \begin{pmatrix} a_1(n) & a_1(n)+1 \\ a_2(n)+1 & a_2(n) \end{pmatrix} x(n-1)\ .$$

There are two possible values for matrices $A(n)$ and the sequence $\{A(n)\}$ alternates between the two values. There are also two possible values for the products $A(n+1)A(n)$. One of them is

$$C = \begin{pmatrix} 3 & 4 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} 8 & 7 \\ 4 & 3 \end{pmatrix}, \quad P(A(n+2p-1)\cdots A(n+1)A(n) = C^p) = \frac{1}{2}\,.$$

Matrix C is a scs1-cyc1 matrix, hence there exists p such that C^p is a rank 1 matrix. As a conclusion, Theorems 6.7.14 or 6.8.5 hold, there is a unique stationary regime for the system.

Now let us consider the same Event Graph with another initial marking : two tokens in place p_2 and no tokens in place p_1. This marking is reachable from the one of Figure 8.1. We consider the new system at the following "virtual" instant : transition t_1 has fired once and transition t_2 has not fired yet. This system is equivalent to the previous one with the firing sequences $\{a_1(n+1), a_2(n),\ n \in \mathbb{N}\}$. Its linear representation is

$$x(n) = \begin{pmatrix} a_1(n+1) & a_1(n+1)+1 \\ a_2(n)+1 & a_2(n) \end{pmatrix} x(n-1)\,.$$

It is easy to check that the products $A(n)A(n-1)\cdots A(0)$ have only two possible forms

$$m \otimes \begin{pmatrix} e & 1 \\ 1 & e \end{pmatrix} \text{ or } m \otimes \begin{pmatrix} 1 & e \\ e & 1 \end{pmatrix},\ m \in \mathbb{N}\,.$$

These matrices are respectively scs2-cyc1 and scs1-cyc2. There is a multiplicity of stationary regimes.

In this example, the sequence $\{(a_1(n), a_2(n))\}$ has not the same distribution as the sequence $\{(a_1(n+1), a_2(n))\}$.

Remark 8.2.4. Note also that in this example, the throughput of the Event Graph (inverse of the eigenvalue of the (max,+) matrix) depends on the initial condition.

Chapitre 9

Task Resource Models and (max,+) Automata

Modèles Tâche Ressource et Automates (max,+)

On montre dans ce chapitre comment l'utilisation d'automates (max,+) permet d'élargir assez sensiblement la classe des systèmes (max,+) linéaires classiques tels qu'ils étaient par exemple présentés au chapitre 1.

On s'intéresse plus spécifiquement à un modèle de ressources partagées, dit modèle Tâche Ressource, représentable sous forme d'automate (max,+). La représentation sous forme d'automate (max,+) permet d'utiliser les résultats du chapitre 6. On propose également des résultats du type optimisation.

Ce chapitre est tiré d'un travail réalisé en commun avec Stéphane Gaubert [72].

We show that a typical class of timed concurrent systems can be modeled as automata with multiplicities in the (max,+) semiring. This representation can be seen as a timed extension of the logical modeling in terms of trace monoids. We briefly discuss the applications of this algebraic modeling to performance evaluation.

9.1 Introduction

Different variations of (stochastic) queuing networks with precedence-based relations between customers have been studied for quite a long time in the performance evaluation community, see [14, 17, 136]. In the combinatorics community on the other hand, concurrent systems are usually modeled in terms of traces —elements of free partially commutative monoids—, see [35, 56]. An equivalent formalism is that of *heaps of pieces* [134].
One of the purposes of this note is to bridge the gap between the two approaches. In the first part of the paper, we establish the relations between the models. An important feature is that execution times of these models can be represented as finite dimensional (max,+) linear dynamical systems. In an essentially equivalent way, they are recognized by automata with multiplicities in the (max,+) semiring. The existence of similar (max,+) models was already noticed in the context of queuing theory [136, 33]. Their analogue for trace monoids seems to be new.
In the second part of the paper, we apply this algebraic modeling to performance evaluation problems. We present asymptotic results on the existence of mean execution time for random schedules, and for optimal and worst schedules. They are obtained by appealing to subadditive arguments borrowed from the theory of random (max,+) matrices [4].
In the third part, we apply the machinery of (max,+) rational series to the exact computation of the asymptotic worst case mean execution time, when the set of admissible schedules is given by a rational language.
At last, some generalizations of Task Resource models are considered (heaps of pieces with arbitrary shapes) for which all the results can be extended. These models provide an algebraic framework to handle scheduling problems.

9.2 Basic Task Resource Model

9.2.1 General presentation

Definition 9.2.1 (Task Resource System). *A (timed)* Task Resource *system is a 4-uple* $\mathcal{T} = (\mathcal{A}, \mathcal{R}, R, h)$ *where:*

- $\mathcal{A}$ *is a finite set whose elements are called* tasks.
- $\mathcal{R}$ *is a finite set whose elements are called* resources.
- $R : \mathcal{A} \to \mathcal{P}(\mathcal{R})$ *gives the subset of resources required by a task. We assume that each task requires at least one resource:* $\forall a \in \mathcal{A}, R(a) \neq \emptyset$.
- $h : \mathcal{A} \to \mathbb{R}^+$ *gives the execution time of a task.*

A length n *schedule* is a sequence of n tasks $a_1, \ldots, a_n$, that we will write as a word[1] $w = a_1 \ldots a_n$. The functioning of the system under the schedule w is as follows.

1. All the resources become initially available at time zero.
2. Task a_i begins as soon as all the required resources $r \in R(a_i)$ used by the earlier tasks $a_j, j < i$, become free, say, at time t_i.
3. Task a_i uses each resource $r \in R(a_i)$ during $h(a_i)$ times units. Thus, resource r is released at time $t_i + h(a_i)$.

The *execution time* or *makespan* of the schedule $w = a_1 \ldots a_n$ is the completion time of the latest task of the schedule (which is not necessarily a_n):

$$y(w) \stackrel{\text{def}}{=} \max_{1 \leqslant i \leqslant n} (t_i + h(a_i)) \ . \tag{9.1}$$

Task Resource systems are intimately related with the classical *trace monoids* that we next define.

Definition 9.2.2. *A* dependence alphabet *is an alphabet $\mathcal{A}$ equipped with a reflexive symmetric relation called* dependence relation*, denoted D, and written graphically —. We denote by I the complement of D (called* independence *relation).*

Definition 9.2.3. *The* trace monoid $\mathbb{M}(\mathcal{A}, D)$ *is the quotient of the free monoid $\mathcal{A}^*$ by the congruence $\sim$ generated by the relations $ab = ba, \forall a\, I\, b$. The elements of $\mathbb{M}(\mathcal{A}, D)$ will be called* traces.

Let alph(w) denote the set of letters appearing in word w. The word $\overline{w} \sim w$ is a *Cartier-Foata normal form* of w [35, 56] if we have a factorization $\overline{w} = u_1 \ldots u_p$, $u_i \in \mathcal{A}^+$, such that:

$$a, b \in \mathrm{alph(u_i)} \Rightarrow \mathrm{a\,I\,b}, \quad \mathrm{a} \in \mathrm{alph(u_i)} \Rightarrow \exists \mathrm{b} \in \mathrm{alph(u_{i-1}), a\,D\,b} \ . \tag{9.2}$$

Such a normal form is unique up to a reordering of the letters inside factors. We shall denote by $\ell(w) = p$ the length (number of factors) of the normal form of w.
With each Task Resource system is associated a dependence relation over the alphabet $\mathcal{A}$; tasks are dependent when they share some resource:

$$a\, D\, b \Leftrightarrow R(a) \cap R(b) \neq \emptyset \ . \tag{9.3}$$

Conversely, starting from an arbitrary trace monoid $\mathbb{M}(\mathcal{A}, D)$, one can build an associated Task Resource system. For example, one can consider $\mathcal{T} = (\mathcal{A}, \mathcal{R}, R, h \equiv 1)$ with $\mathcal{R} = \{\{a, b\} \mid a\, D\, b\}$ and $R(a) = \{r \in \mathcal{R} \mid a \in r\}$. The problem of finding a system $\mathcal{T}$ such that the cardinality of $\mathcal{R}$ is minimal is considered in Appendix.

[1] We recall the following usual notation. Given a finite set (alphabet) $\mathcal{A}$, we denote by $\mathcal{A}^n$ the set of words of length n on $\mathcal{A}$. We denote by $\mathcal{A}^*$ the free monoid on $\mathcal{A}$, that is, the set of finite words equipped with concatenation. The unit (empty word) will be denoted e. We denote by $\mathcal{A}^+ = \mathcal{A}^* \setminus \{e\}$ the free semigroup on $\mathcal{A}$. The length of the word w will be denoted $|w|$. We shall write $|w|_a$ for the number of occurrences of a given letter a in w.

Proposition 9.2.4. (i) *When $h \equiv 1$, $y(w) = \ell(w)$: the makespan is equal to the length of the Cartier-Foata normal form of w.* (ii) *For general execution times h,*

$$y(w) = \max \sum_{j=1}^{p} h(a_{i_j}) \ , \tag{9.4}$$

where the max is taken over the subwords $a_{i_1} \dots a_{i_p}$ of $w = a_1 \dots a_n$, composed of consecutive dependent letters (i.e. $a_{i_j} \, D \, a_{i_{j+1}}$).

The first assertion is classical [36]. It implies in particular that the makespan of Task-Resource systems with $h \equiv 1$ can be represented in a more intrinsic way in terms of trace monoid. The second one can easily be proved by elementary means, or deduced from the (max,+)-linear representation given below. It provides an alternative formula for (9.1).

Example 9.2.5.
For the sequential dependence alphabet $a \, D \, b$, we have $y(w) = h(a)|w|_a + h(b)|w|_b$. For the purely parallel dependence alphabet $a \, I \, b$, we have $y(w) = \max(h(a)|w|_a, h(b)|w|_b)$.

Example 9.2.6. [Ring Network] Consider a ring shaped communication network with k stations $\mathcal{R} = \{r_1, \dots, r_k\}$. Messages can be sent between neighbor stations. The possible messages are $\mathcal{A} = \{a_1, \dots, a_k\}$ where a_i corresponds to a communication between r_i and r_{i+1} (with the convention $k+1 = 1$). Therefore, we have $R(a_i) = \{r_i, r_{i+1}\}$. This system can also be viewed as a variant of the classical dining philosophers model [57] (replace stations by chopsticks, messages by philosophers). E.g., for $k = 5$, $y(a_1a_2a_4a_1a_5) = \max(2h(a_1) + h(a_2) + h(a_5), h(a_4) + h(a_5))$ (direct application of 9.2.4,(ii) since the maximal dependent subwords taken from $a_1a_2a_4a_1a_5$ are $a_1a_2a_1a_5$ and a_4a_5).

9.2.2 Linear representation over the (max,+) semiring

Definition 9.2.7. *The* (max,+) semiring $\mathbb{R}_{\max}$ *is the set* $\mathbb{R} \cup \{-\infty\}$, *equipped with* max, *written additively (i.e. $a \oplus b = \max(a, b)$) and the usual sum, written multiplicatively (i.e. $a \otimes b = a + b$). We write $\varepsilon = -\infty$ for the zero element, and $e = 0$ for the unit element.*

We shall use throughout the paper the matrix and vector operations induced by the semiring structure[2]. The identity matrix ($\mathrm{I_{ii}} = \mathrm{e}, \mathrm{I_{ij}} = \varepsilon, \mathrm{i} \neq \mathrm{j}$) with entries indexed by X will be denoted by $\mathrm{I_X}$. The row vector with entries indexed by X and all equal to e will be denoted by e_X. We denote by $\|M\| = \bigoplus_{ij} M_{ij}$ (resp. $\|v\| = \bigoplus_i v_i$) the (max,+) *norm* of a matrix M (vector v).

A (max,+) *automaton*[3] of dimension k over the alphabet $\mathcal{A}$ is a triple $(\alpha, \mathcal{M}, \beta)$, where $\alpha \in \mathbb{R}_{\max}^{1 \times k}$, $\beta \in \mathbb{R}_{\max}^{k \times 1}$, and $\mathcal{M}$ is a morphism from $\mathcal{A}^*$ to the multiplicative monoid of matrices $\mathbb{R}_{\max}^{k \times k}$. A map $y : \mathcal{A}^* \to \mathbb{R}_{\max}$ is *recognizable* if there is an automaton such that $y(w) = \alpha \mathcal{M}(w) \beta$.

[2] I.e. for matrices A, B of appropriate sizes, $(A \oplus B)_{ij} = A_{ij} \oplus B_{ij} = \max(A_{ij}, B_{ij})$, $(A \otimes B)_{ij} = \bigoplus_k A_{ik} \otimes B_{kj} = \max_k(A_{ik} + B_{kj})$, and for a scalar a, $(a \otimes A)_{ij} = a \otimes A_{ij} = a + A_{ij}$. We will abbreviate $A \otimes B$ to AB as usual.

[3] This is a specialization to the $\mathbb{R}_{\max}$ case of the notion of automaton with multiplicities over a semiring (or equivalently, of recognizable series over a semiring). See [60, 21].

In a spirit closer to discrete event systems theory, automata may be seen as (max,+) linear systems whose dynamics is indexed by letters. Indeed, introducing the "state vector" $x(w) \stackrel{\text{def}}{=} \alpha\mathcal{M}(w) \in \mathbb{R}_{\max}^{1\times k}$, we get

$$x(e)=\alpha,\quad x(wa) = x(w)\mathcal{M}(a),\ y(w)=x(w)\beta\ ,\quad \text{or} \tag{9.5}$$

$$y(a_1\dots a_n) = \alpha\mathcal{M}(a_1)\dots\mathcal{M}(a_n)\beta\ . \tag{9.6}$$

Definition 9.2.8 (Task & Resource Daters). *A* dater *over the alphabet $\mathcal{A}$ is a scalar map $\mathcal{A}^* \to \mathbb{R}\cup\{-\infty\}$. With each task $a\in\mathcal{A}$ is associated a* task dater x_a*: $x_a(w)$ gives the time of completion of the last task of type a in the schedule w. With each resource $r\in\mathcal{R}$ is associated a* resource dater x_r*: $x_r(w)$ gives the last instant of release of the resource r under the schedule w. We shall denote by $x_\mathcal{A}$ and $x_\mathcal{R}$ the vectors of task and resource daters.*

Note the important duality relations

$$x_a(w) = \bigoplus_{r\in R(a)} x_r(w),\ \ x_r(w) = \bigoplus_{a\in R^{-1}(r)} x_a(w)\ . \tag{9.7}$$

We identify each subset $R(a)$ with a boolean matrix of size $|\mathcal{R}|\times|\mathcal{A}|$ denoted $\mathcal{I}(a)$.

$$\forall a\in\mathcal{A}, \mathcal{I}(a)_{rb} = \begin{cases} e & \text{if } r\in R(a) \text{ and } b=a \\ \varepsilon & \text{otherwise}\ . \end{cases}$$

We define the following matrices:

$$\forall a\in\mathcal{A},\quad \mathcal{M}_\mathcal{R}(a) = \mathrm{I}_\mathcal{R} \oplus \mathrm{h(a)}\mathcal{I}(\mathrm{a})\mathcal{I}(\mathrm{a})^{\mathrm{T}}\ , \tag{9.8}$$

$$\mathcal{M}_\mathcal{A}(a) = \mathrm{I}_\mathcal{A} \oplus \mathrm{h(a)}\left(\bigoplus_{\mathrm{b}} \mathcal{I}(\mathrm{b})^{\mathrm{T}}\right)\mathcal{I}(\mathrm{a})\ , \tag{9.9}$$

or more explicitly

$$\mathcal{M}_\mathcal{R}(a)_{rs} = \begin{cases} e & \text{if } r=s, s\notin R(a), \\ h(a) & \text{if } r\in R(a), s\in R(a), \\ \varepsilon & \text{otherwise.} \end{cases} \tag{9.10}$$

$$\mathcal{M}_\mathcal{A}(a)_{bc} = \begin{cases} e & \text{if } a\neq(b=c), \\ h(a) & \text{if } a=c, bDc, \\ \varepsilon & \text{otherwise.} \end{cases} \tag{9.11}$$

We extend $\mathcal{M}_\mathcal{A}$ (resp. $\mathcal{M}_\mathcal{R}$) to a morphism $\mathcal{A}^*\to\mathbb{R}_{\max}^{\mathcal{A}\times\mathcal{A}}$ (resp. $\mathcal{A}^*\to\mathbb{R}_{\max}^{\mathcal{R}\times\mathcal{R}}$).

Theorem 9.2.9. *The dater functions of task resource systems admit the following linear representations over the (max,+) semiring:*

$$x_\mathcal{R}(wa) = x_\mathcal{R}(w)\mathcal{M}_\mathcal{R}(a),\quad x_\mathcal{R}(e)=e_\mathcal{R}\ , \tag{9.12}$$

$$x_\mathcal{A}(wa) = x_\mathcal{A}(w)\mathcal{M}_\mathcal{A}(a),\quad x_\mathcal{A}(e)=e_\mathcal{A}\ , \tag{9.13}$$

$$y(w) = \|x_\mathcal{A}(w)\| = \|x_\mathcal{R}(w)\| = \|\mathcal{M}_\mathcal{A}(w)\| = \|\mathcal{M}_\mathcal{R}(w)\|\ . \tag{9.14}$$

In other words, y is recognized both by the *resource automaton* $(e_{\mathcal{R}}, \mathcal{M}_{\mathcal{R}}, e_{\mathcal{R}}^T)$ and by the *task automaton* $(e_{\mathcal{A}}, \mathcal{M}_{\mathcal{A}}, e_{\mathcal{A}}^T)$.

Proof. We have

$$x_a(wb) = \begin{cases} x_a(w) & \text{if } a \neq b, \\ \max_{r \in R(a)} x_r(w) + h(a) & \text{if } a = b, \end{cases} \tag{9.15}$$

$$x_r(e) = x_a(e) = e \ . \tag{9.16}$$

These relations are a simple translation of the functioning of the system, as described after Definition 9.2.1 (items 1,2,3). Eliminating x_r in (9.15) using (9.7), we get the *task equation*

$$x_a(wb) = \begin{cases} x_a(w) & \text{if } a \neq b \\ \max_{cDa} x_c(w) + h(a) & \text{if } a = b. \end{cases} \tag{9.17}$$

Dually, it is not difficult to obtain the *resource equation*

$$x_r(wa) = \begin{cases} x_r(w) & \text{if } R(a) \not\ni r \\ \max_{s \in R(a)} x_s(w) + h(a) & \text{if } R(a) \ni r. \end{cases} \tag{9.18}$$

Rewriting (9.17) and (9.18) with the semiring notations, we get (9.12),(9.13). □

Remark 9.2.10. Note that the duality is not perfect in a task resource model. In both Equations (9.12) and (9.13), the dynamic of the system is driven by a word w which is a sequence of tasks. A sequence of resources would have no meaning.

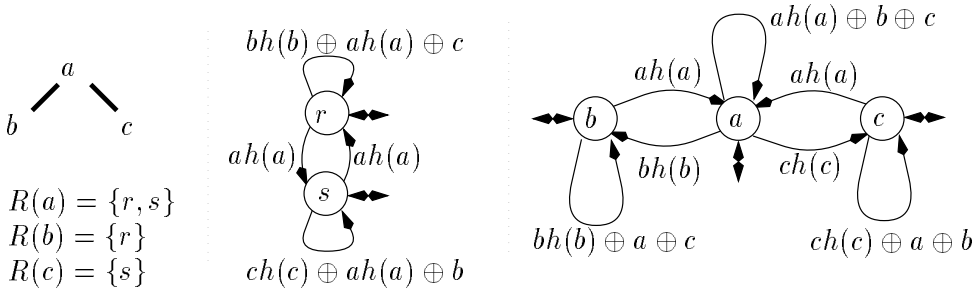


Figure 9.1: Task and Resource Automata for b—a—c.

Example 9.2.11. We consider a Task Resource model. Let b—a—c be its dependence alphabet. In Fig. 9.1, we have represented[4] the resource automaton $(e_{\mathcal{R}}, \mathcal{M}_{\mathcal{R}}, e_{\mathcal{R}}^T)$ and the task

[4] An automaton $(\alpha, \mathcal{M}, \beta)$ of dimension k over an alphabet $\mathcal{A}$ is usually represented as a graph with nodes $1, \ldots, k$, and three kinds of labeled and weighted arcs. There is an *internal arc* $i \to j$ with label $a \in \mathcal{A}$ and weight

automaton $(e_{\mathcal{A}}, \mathcal{M}_{\mathcal{A}}, e_{\mathcal{A}}^T)$ associated with the dependence alphabet b—a—c. The matrices associated with the resource automaton are:

$$\mathcal{M}_{\mathcal{R}}(a) = \begin{bmatrix} h(a) & h(a) \\ h(a) & h(a) \end{bmatrix}, \mathcal{M}_{\mathcal{R}}(b) = \begin{bmatrix} h(b) & \varepsilon \\ \varepsilon & e \end{bmatrix}, \ \mathcal{M}_{\mathcal{R}}(c) = \begin{bmatrix} e & \varepsilon \\ \varepsilon & h(c) \end{bmatrix} \ .$$

The makespan $y(w)$ is equal to the maximal weight of a path labeled w between two arbitrary nodes of the graph. E.g., $y(cba) = \max(h(c) + h(a), h(b) + h(a))$.

Example 9.2.12. We consider a Task Resource model with $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$ and $\mathcal{R} = \{r_1, r_2, r_3, r_4\}$. Let $R(a_1) = (r_1, r_2)$, $R(a_2) = (r_2, r_4)$, $R(a_3) = (r_1, r_2, r_3, r_4)$, $R(a_4) = (r_3)$, $R(a_5)$=(r_3) and $h(a_i) = 1, i = \{1, \ldots, 4\}, h(a_5) = 3$.
Let us give the matrices associated with letter a_2. We have :

$$M_{\mathcal{R}}(a_2) = \begin{pmatrix} e & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 1 & \varepsilon & 1 \\ \varepsilon & \varepsilon & e & \varepsilon \\ \varepsilon & 1 & \varepsilon & 1 \end{pmatrix}, M_{\mathcal{A}}(a_2) = \begin{pmatrix} e & \varepsilon & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 1 & \varepsilon & 1 & \varepsilon \\ \varepsilon & \varepsilon & e & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & e & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & e \end{pmatrix} .$$

For letter a_5, we have $M_{\mathcal{R}}(a_5)_{33} = 3, M_{\mathcal{R}}(a_5)_{ij} = \varepsilon, \forall (i,j) \neq (3,3)$ and $M_{\mathcal{A}}(a_5)_{5j} = 3, j = \{3,4,5\}, M_{\mathcal{A}}(a_5)_{ij} = \varepsilon$, otherwise.

9.2.3 Interpretation in terms of hypergraphs

Task resource models can be introduced as hypergraphs. It is the approach used for example by Vincent [137] in a work concurrent to this one. An hypergraph is a direct generalization of a non-oriented graph. It has a set of vertices and a set of edges which are subsets of vertices (instead of couples of vertices). Let us define $R(\mathcal{A}) = \{R(a),\ a \in \mathcal{A}\}$. The couple $(\mathcal{R}, R(\mathcal{A}))$ defines an hypergraph. The set of vertices is $\mathcal{R}$ and the set of edges is $R(\mathcal{A})$. Let us define the function $A : \mathcal{R} \to \mathcal{P}(\mathcal{A})$ by $A(r) = \{a \in \mathcal{A} \mid r \in R(a)\}$ and let us define $A(\mathcal{R}) = \{A(r), r \in \mathcal{R}\}$. The couple $(\mathcal{A}, A(\mathcal{R}))$ is an hypergraph. These two hypergraphs are dual, see Berge [20].
The dual (max,+) equations describing a task resource model, see Equations (9.12) and (9.13) correspond to this duality.

9.2.4 Interpretation in terms of heaps of pieces

There is a useful geometrical interpretation of Task Resource Models in terms of *heaps of pieces*. This interpretation was first noticed by Viennot for trace monoids. The reader is referred to [134] for a more formal presentation. Imagine an horizontal axis with as many slots as resources. With each letter a is associated a *piece*, i.e. a solid "rectangle" occupying the slots $r \in R(a)$, with height $h(a)$. The heap associated with the word $w = a_1 \ldots a_n$ is built by piling up the

$\mathcal{M}(a)_{ij}$ whenever $\mathcal{M}(a)_{ij} = t \neq \varepsilon$. We will write $x \xrightarrow{at} y$ but we omit the unit valuations (when $t = e$). When there are two arcs $x \to y$ with respective labels a, b and weights t, t', we shall write $x \xrightarrow{at \oplus bt'} y$ as a shorthand for the two arcs $x \xrightarrow{at} y$, $x \xrightarrow{bt'} y$. There is an *input arc* at node i with weight α_i, whenever $\alpha_i \neq \varepsilon$. *Output arcs* are obtained in a dual way from β.

pieces $a_1, \ldots, a_n$, in this order. The makespan $y(w)$ coincides with the height of the heap. The vector $x_{\mathcal{R}}(w) = e_{\mathcal{R}}\mathcal{M}_{\mathcal{R}}(w)$ can be interpreted as the upper contour of the heap. Adding one piece above the heap amounts to right multiplication by the corresponding matrix.

Example 9.2.13. Consider the Task Resource model of Example 9.2.12. We have represented, in Figure 9.2, the heap associated with the word $w = a_1a_1a_5a_2a_3a_1a_2a_4$. Piece a_2 is an example of a solid but not connected piece.

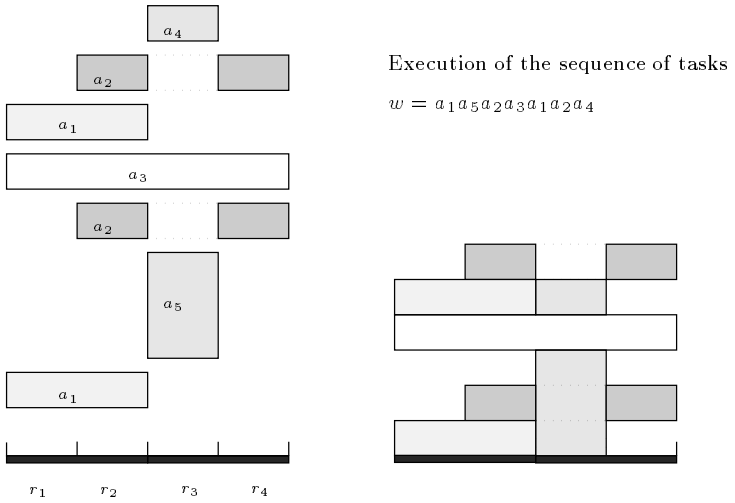


Figure 9.2: Heap of pieces for a Task Resource model.

9.3 Performance Evaluation

9.3.1 Stochastic case

The simplest[5] stochastic extension of task resource systems arises when the sequence of tasks is given by a sequence of random variables $a(n) \in \mathcal{A}$: we get the random schedule $w_n = a(1) \ldots a(n)$, and consider the asymptotics of $y(w_n), x(w_n)$, that we shall shorten to $y(n), x(n)$. For stochastic Task Resource models, we propose two types of asymptotic results.

1. First order limits or mean execution times $x(n)_i/n$.
2. Second order limits or asymptotics of relative delays $x(n)_i - x(n)_j$ (e.g. differences of last occupation times of the different resources).

Second order quantities are best defined in terms of (max,+) projective space. The (max,+) projective space $\mathbb{PR}^k_{\max}$ is the quotient of $\mathbb{R}^k$ by the parallelism relation $x \simeq y \Leftrightarrow \exists \lambda \in \mathbb{R}$, $x = \lambda y$. We write $\pi : \mathbb{R}^k_{\max} \to \mathbb{PR}^k_{\max}$ the canonical projection. The relative delays $x(n)_i - x(n)_j$ can be computed from $\pi x(n)$. Geometrically, $\pi x(n)$ corresponds to *the upper shape of the heap*

[5] In order to simplify the presentation, we shall not consider more general cases with random initial conditions, random executions times and random arrival times, which can be dealt with along the same lines.

(the quotient by $\simeq$ identifies two heaps with the same upper contour but different heights, see Fig. 9.3).
We assume that the random variables $a(n)$ are defined on a common probability space $(\Omega, \mathcal{F}, P)$, equipped with a stationary and ergodic shift θ. We consider a *connected* Task Resource system, i.e. such that the graph of the dependence relation is connected (if it is not the case, the theorem has to be applied to each connected sub-system).

Theorem 9.3.1. *Let $\{a(n), n \in \mathbb{N}\}$ be a stationary and ergodic sequence (i.e. $a(n+1,\omega) = a(n,\theta(\omega))$) of integrable random variables, such that $\forall b \in \mathcal{A}, P(a(1) = b) > 0$.*

1. *There exists a constant $\lambda_E \in \mathbb{R}$ (*stochastic Lyapunov exponent*) such that, $\forall i \in \mathcal{A} \cup \mathcal{R}$,*

$$\lim_n \frac{x(n)_i}{n} = \lim_n E\left(\frac{x(n)_i}{n}\right) = \lambda_E \quad P - \text{a.s.} \tag{9.19}$$

2. *Moreover, if the sequence $\{a(n), n \in \mathbb{N}\}$ is i.i.d. then the random variable $\pi x(n)$ converges in total variation to a unique stationary distribution.*

Proof. In order to prove point 1, the main tool is the subadditivity of the sequence $\{y(w) = \|x(w)\|\}$, more precisely:

$$\forall w_1, w_2 \in \mathcal{A}^*, \;\; y(w_1 w_2) \leqslant y(w_1) + y(w_2) \; . \tag{9.20}$$

This property enables to apply Kingman's subadditive ergodic theorem, see [4]. More generally, this result is just a special case of a general theorem on homogeneous and monotone operators, see [136] or Theorem 10.3.5.
We show point 2 for the resource dater $x_{\mathcal{R}}(w) = e_{\mathcal{R}} \mathcal{M}_{\mathcal{R}}(w)$ (the behavior of $x_{\mathcal{A}}$ can be deduced easily from that of $x_{\mathcal{R}}$ by appealing to (9.7)). The following necessary and sufficient condition of existence and uniqueness of a stationary distribution for $\pi x_{\mathcal{R}}(w(n))$ is stated in Chapter 6:

There is a word w such that the matrix $\mathcal{M}_{\mathcal{R}}(w)$ is of rank one, with non-ε entries.

The matrix $\mathcal{M}_{\mathcal{R}}(w)$ constitutes a regeneration pattern for the model. Indeed, the rank one condition is equivalent to a forgetting of the initial condition.

$$\forall x_0, x_0', \;\; \pi(x_0 \mathcal{M}_{\mathcal{R}}(w)) = \pi(x_0' \mathcal{M}_{\mathcal{R}}(w)) \; . \tag{9.21}$$

This pattern enables us to use regeneration theory to obtain stability of the model. The existence of the pattern is guaranteed by the following lemma.

Lemma 9.3.2. *Let $w = a_1 \ldots a_n$ be a path in the graph of the dependence relation ($a_i \, D \, a_{i+1}$), visiting all the nodes. Let $\tilde{w} = a_n \ldots a_1$ denote the mirror image of w. The matrix $\mathcal{M}_{\mathcal{R}}(w\tilde{w})$ is of rank one with non-ε entries.*

Rather than proving formally the result (which can be done using representation (9.8), (9.12) and the fact that $\mathcal{I}(a)$ has rank one), we provide a geometrical justification using heaps of pieces. Condition (9.21) is equivalent to the following: *the upper shape of the heap is independent of the shape of the ground* (which corresponds to the initial condition x_0). The property $a_i \, D \, a_{i+1}$ of the word $w\tilde{w}$ means that the heap is staircase shaped. It implies condition (9.21), see the example below.

Example 9.3.3. Consider the ring model of Example 9.2.6 with $k = 4$ and $h \equiv 1$. We have represented in Fig. 9.3.(I), the heap associated with the word $a_1a_2a_3a_4a_4a_3a_2a_1$.

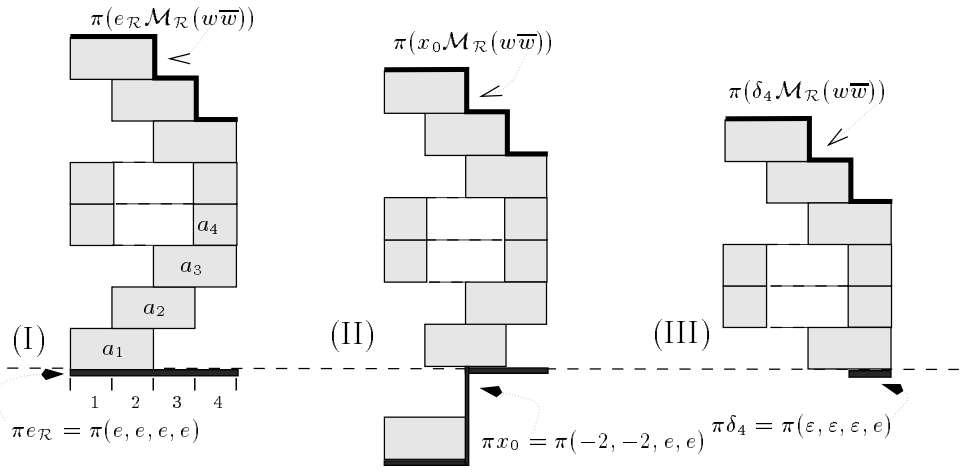


Figure 9.3: Heaps of pieces for a ring model.

The upper shape is independent of the shape of the ground as illustrated in the different heaps (I),(II),(III) shown on Fig. 9.3 (corresponding to the respective initial conditions $e_{\mathcal{R}}$, $(-2,-2,e,e)$ and $(\varepsilon,\varepsilon,\varepsilon,e)$).

□

Remark 9.3.4. A result analog to Theorem 9.3.1, point 2. was proved by Saheb [126] for trace monoids, using a Markovian argument. The advantage of the method presented here is that it can be applied to the various extensions mentioned in footnote 5.

When the sequence $\{a(n), n \in \mathbb{N}\}$ is only stationary and ergodic, the necessary and sufficient condition of stability for $\pi x_{\mathcal{R}}(w(n))$ is (see Chapter 6) :

There is a word w of length m such that the matrix $\mathcal{M}_{\mathcal{R}}(w)$ is of rank one with non-ε entries, and $P\{a(1)\ldots a(m) = w\} > 0$.

Under stationary and ergodic assumptions, Theorem 9.3.1, point 2. is not true. Indeed, the mirror words $w\tilde{w}$ defined in Lemma 9.3.2 might appear with probability 0. Here is an example.

Example 9.3.5. Consider the ring model of Example 9.2.6 with $k = 5$ and $h \equiv 1$. We consider a probability space (Ω, P) with $\Omega = \{\omega_1, \ldots, \omega_6\}$ and $P = \{\frac{1}{6}, \ldots, \frac{1}{6}\}$. Let θ be the stationary

and ergodic shift defined by : $\forall i,\ \theta(\omega_i) = \omega_{i+1}$ [6]. We consider the sequence of random variables :

$$\begin{aligned}\{a(n,\omega_1), n \in \mathbb{N}\} &= (a_2a_4a_4a_1a_1a_3)(a_2a_4a_4a_1a_1a_3)\cdots \\ \{a(n,\omega_2), n \in \mathbb{N}\} &= (a_4a_4a_1a_1a_3a_2)(a_4a_4a_1a_1a_3a_2)\cdots,\ \ \ldots\end{aligned}$$

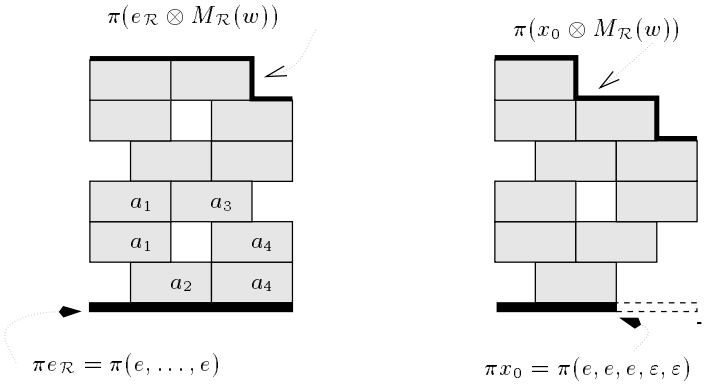


Figure 9.4: Heap of pieces associated with the word $w = (a_2a_4a_4a_1a_1a_3)^2$.

We have represented the heaps of pieces associated with the word $w = (a_2a_4a_4a_1a_1a_3)^2$ for two different initial conditions (Figure 9.4). Their upper contour is not the same, hence matrix $M_{\mathcal{R}}(w)$ is not of rank 1. It is easy to verify that it is the case for all the matrices which appear in this model.

9.3.2 Optimal case and worst case

Given a language $L \subset \mathcal{A}^*$ describing the set of admissible schedules, a natural problem consists in finding an admissible schedule of length n with minimal or maximal makespan. The following theorem shows the existence of an asymptotic mean execution time, under optimal or worst case schedules. It can be seen as a (weak) analogue for optimization problems of the first order ergodic theorem 9.3.1,1.

Theorem 9.3.6. *1. For a language L such that $L^2 \subset L$, the following limit (optimal Lyapunov exponent) exists*

$$\lambda_{\min}(L) \stackrel{\text{def}}{=} \lim_{n\to\infty,\ \mathcal{A}^n\cap L\neq\emptyset} \min_{w\in\mathcal{A}^n\cap L} \frac{y(w)}{n} = \inf_{w\in L} \frac{y(w)}{|w|}\ . \tag{9.22}$$

2. For a bifix *language L (such that $uv \in L \Rightarrow u,v \in L$), the following limit (worst Lyapunov exponent) exists*

$$\lambda_{\max}(L) \stackrel{\text{def}}{=} \lim_{n\to\infty} \max_{w\in\mathcal{A}^n\cap L} \frac{y(w)}{n} = \inf_{n\geqslant 1} \max_{w\in\mathcal{A}^n\cap L} \frac{y(w)}{n}\ . \tag{9.23}$$

Proof. Let $m_n = \inf_{w\in\mathcal{A}^n\cap L} y(w)$. Since $L^2 \subset L$, $w \in L\cap\mathcal{A}^n, z \in L\cap\mathcal{A}^p \Rightarrow wz \in L\cap A^{n+p}$. Using the subadditivity property (9.20), we get $m_{n+p} \leqslant m_n + m_p$, from which (9.22) readily follows. The argument for $\lambda_{\max}$ is similar. □

The assumption that $L^2 \subset L$ for the optimal case is practically reasonable. For instance, for usual scheduling problems, it is natural to impose a fixed proportion of the different tasks, i.e. $L = \{w \mid |w|_a = r_a|w|\}$, for some fixed $r_a \in \mathbb{R}^+, \sum_a r_a = 1$. Such a language satisfies $L^2 \subset L$. The restriction to bifix languages for the worst case behavior is an artefact due to the subadditive argument.
The following theorem shows that the worst case performance can be exactly computed for the subclass of rational schedule languages. The reader is referred to [21] for the notation concerning series.

Theorem 9.3.7.
Consider the generating series *of the worst case behavior,* $z = \bigoplus_{n\in\mathbb{N}} z_n x^n \in \mathbb{R}_{\max}[[x]]$, *where* $z_n = \sup_{w\in\mathcal{A}^n\cap L} y(w)$. *If the admissible language* L *is rational, the series* z *is rational.*

Proof. Let charL $\in \mathbb{R}_{\max}\langle\langle\mathcal{A}\rangle\rangle$ denote the characteristic series[6] of the language L. Then, charL is rational. Introduce the morphism $\varphi : \mathbb{R}_{\max}\langle\langle\mathcal{A}\rangle\rangle \to \mathbb{R}_{\max}[[x]]$ such that $\forall a, \varphi(a) = x$. Recall that the *Hadamard product* of series is defined by $(s\odot t)(w) = s(w)t(w)$. Since rational series are closed under alphabetical morphisms and Hadamard product, $z = \varphi(\text{charL}\odot \text{y}) \in \mathbb{R}_{\max}[[\text{x}]]$ is rational. □

Corollary 9.3.8. *Let* α, μ, β *denote a trim linear representation of* charL. *Then,*

$$\limsup_n \frac{z_n}{n} = \rho(A), \quad A = \bigoplus_{a\in\mathcal{A}} \mu(a) \otimes^t \mathcal{M}_{\mathcal{R}}(a) \ , \tag{9.24}$$

where ρ *denotes the (max,+) maximal eigenvalue and* $\otimes^t$ *the tensor product of matrices.*

This is an immediate consequence of the (max,+) spectral theorem, together with the fact [60, 21] that charL$\odot$y is recognized by the tensor product of the representations (α, μ, β), $(e_{\mathcal{R}}, \mathcal{M}_{\mathcal{R}}, e_{\mathcal{R}}^T)$ (see [69, §3.2] for details).

Remark 9.3.9. More generally, Theorem 9.3.7 holds for an algebraic (=context-free) language L and not only for a rational one. Indeed, it is an easy extension[7] of Parikh theorem [47] that *algebraic series in several commuting indeterminates, with coefficients in* $\mathbb{R}_{\max}$, *are rational.* Since algebraic series are closed by Hadamard product with recognizable series and alphabetical morphism, the above proof shows that, when L is algebraic, the series $z = \varphi(\text{charL}\odot \text{y})$ is algebraic, hence rational. This shows that *the generating series* z *of the worst case behavior of an algebraic language* L *is rational.* In this case, the effective computation of z, along the lines of [47, Ch. XI] is less immediate, since it requires solving (max,+) commutative rational equations.

[6] The coefficient of charL at w is equal to e if $w \in L$, ε otherwise.

[7] By algebraic series, we mean *constructive* algebraic series as defined in [63]. The argument given in [47, Ch. XI] can be adapted to algebraic series in commuting indeterminates with coefficients in commutative idempotent semirings.

Example 9.3.10. Consider the dependence alphabet b—a—c, together with the set of admissible schedules $L = (a \oplus bc^*b)^*$. Its characteristic series is recognized by

$$\alpha = [e, \varepsilon], \beta = [e, \varepsilon]^T, \mu(a) = \begin{bmatrix} e & \varepsilon \\ \varepsilon & \varepsilon \end{bmatrix}, \ \mu(b) = \begin{bmatrix} \varepsilon & e \\ e & \varepsilon \end{bmatrix}, \ \mu(c) = \begin{bmatrix} \varepsilon & \varepsilon \\ \varepsilon & e \end{bmatrix} \ .$$

We get from Ex. 9.2.11 and (9.24),

$$A = \begin{bmatrix} h(a) & h(a) & h(b) & \varepsilon \\ h(a) & h(a) & \varepsilon & e \\ h(b) & \varepsilon & e & \varepsilon \\ \varepsilon & e & \varepsilon & h(c) \end{bmatrix} \ , \quad \rho(A) = h(a) \oplus h(c) \oplus h(b) \ ,$$

where $\rho(A)$ is obtained from its characterization as maximal mean weight of the circuits of A [8]. Note that the different terms in $\rho(A)$ are attained asymptotically for the sequences of schedules $a^n, n \in \mathbb{N}$, $bc^nb, n \in \mathbb{N}$, $b^{2n}, n \in \mathbb{N}$ (whose periodic parts correspond to circuits of A).

Remark 9.3.11. Cérin and Petit [36] study the absolute worst case behavior defined by $\overline{\lambda}_{\max} \stackrel{\text{def}}{=} \sup_{w \in L} |w|^{-1} \times y(w)$. This can be obtained along the same lines:

$$\overline{\lambda}_{\max} = \rho(A) \oplus \bigoplus_{1 \leqslant i \leqslant \dim A} cA^ib \ , \tag{9.25}$$

where $c = \alpha \otimes^t e_{\mathcal{R}}, b = \beta \otimes^t e_{\mathcal{R}}^T$. These quantities can be computed in $O((\dim A)^3)$ steps (using Karp algorithm [8] for $\rho(A)$). Observe that the dual quantity $\inf_{w \in L} y(w)/|w|$ treated in [36] cannot be obtained by such simple arguments due to its "min-max" structure.

9.4 Extensions of Task Resource Models

9.4.1 Tetris game

Tetris is a famous electronic game. It consists in pieces of different forms (tetraminos in the original version). They fall from above one after the other. They stop as soon as they meet another piece. In fact, a game of Tetris looks exactly like Figure 9.2, except for the form of the pieces.

Definition 9.4.1. *We define a Tetris type model (or generalized Task Resource model) as a 5-uple $(\mathcal{A}, \mathcal{R}, R, B_{\mathcal{A}}, T_{\mathcal{A}})$, where :*

- *$\mathcal{A}$ is a finite set of pieces (or tasks).*
- *$\mathcal{R}$ is a finite set of slots (or resources).*
- *$R : \ \mathcal{A} \longrightarrow \mathcal{P}(\mathcal{R})$ gives the subset of slots covered by a piece.*
- *$B_a : R(a) \longrightarrow \mathbb{R}_+^*$, $a \in \mathcal{A}$. The letter B stands for Bottom of the piece. By convention, the map B_a is chosen such that $\max_{r \in R(a)} B_a(r) = 0$.*
- *$T_a : R(a) \longrightarrow \mathbb{R}_+^*$, $a \in \mathcal{A}$. The letter T stands for Top of the piece. The map T_a satisfies the relation $T_a \geqslant B_a$.*

Remark 9.4.2. The basic Task Resource model (Definition 9.2.1) is a Tetris type model where each piece a verifies $\forall r \in R(a), B_a(r) = 0$ and $\forall r \in R(a), T_a(r) = h(a)$.

We extend the definition of task and resource daters, Def. 9.2.8, to Tetris type models. Let us define

$$\mathcal{M}_{\mathcal{R}}(a)_{rs} = \begin{cases} e & \text{if } r = s, s \notin R(a), \\ T_a(s) - B_a(r) & \text{if } r \in R(a), s \in R(a), \\ \varepsilon & \text{otherwise.} \end{cases} \tag{9.26}$$

We extend $\mathcal{M}_{\mathcal{R}}$ to a morphism $\mathcal{A}^* \to \mathbb{R}_{\max}^{\mathcal{R}\times\mathcal{R}}$. Theorem 9.2.9 still applies, i.e. we have :

$$x_{\mathcal{R}}(wa) = x_{\mathcal{R}}(w)\mathcal{M}_{\mathcal{R}}(a), \quad x_{\mathcal{R}}(e) = e_{\mathcal{R}} \ . \tag{9.27}$$

It is also possible to define a dual (max,+) representation for Tetris models. However the expressions for the matrices $\mathcal{M}_{\mathcal{A}}(a)$ are not as simple as in Equation (9.26).

Theorem 9.4.3.
All the results of §9.3 can be extended to Tetris type models.

Example 9.4.4. We consider, as an example, the tetramino of Figure 9.5.

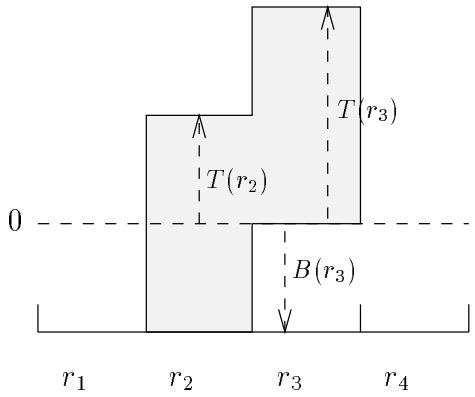


Figure 9.5: Piece of a Tetris type model.

This tetramino a is defined by $R(a) = (r_2, r_3), B_a(r_2) = 0, B_a(r_3) = 1, T_a(r_2) = 2, T_a(r_3) = 3$. The matrix associated with piece a is :

$$M_{\mathcal{R}}(a) = \begin{pmatrix} e & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 2 & 3 & \varepsilon \\ \varepsilon & 1 & 2 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & e \end{pmatrix} .$$

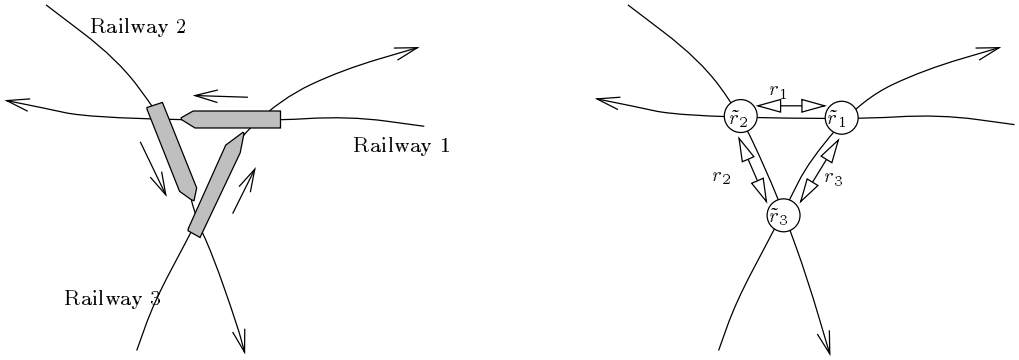


Figure 9.6: A simplified model of the Gagny's triangle.

9.4.2 Gagny's triangle

Gagny is a town near Paris. The railway network in Gagny has a structure which is presented in Figure 9.6.
It consists of three tracks. A train using track i can leave the triangle if and only if track $(i+1)$ [3] is empty. Otherwise, it has to wait until the train on track $(i+1)$ [3] leaves. An inter-blocking is possible if there is a train on each track at the same time. This event has to be avoided. One easy way to avoid deadlocks is the following : as soon as a train enters track i, a light switches to red at track $(i+1)$ [3] preventing any train to enter it (the red light has no effect on a train already on track $(i+1)$). When the train leaves track i, the light at track $(i+1)$ switches back to green. With this control, there is a maximum of two trains at the same time in the triangle.

It might not be very intuitive but this network can be represented as a Tetris type model. Track i is modelled as consisting of two portions, r_i and $\tilde{r}_i$. The set of resources is $\{\tilde{r}_i, r_i,\ i = 1,2,3\}$. The set of tasks is $\{a_i, i = 1,2,3\}$, where a_i corresponds to the passage of a train on track i. We consider a simple model where the passage of a train can be decomposed in two stages.

1. The train enters track i. It requires the railway portions r_i and $\tilde{r}_i$ during t_i units of time.
2. The train leaves track i. It requires the railway portions r_i and $\tilde{r}_{i+1}$ during η_i units of time.

Once it enters the triangle, say at instant t, a train a_i blocks three resources : tracks r_i and $\tilde{r}_i$ but also track $\tilde{r}_{i+1}$ (because of the red light). By blocking, we mean that the tracks can not be used by the next trains. However in a first time, only the tracks r_i and $\tilde{r}_i$ are used by train a_i. At instant $t+t_i$, the train is ready to leave track $\tilde{r}_i$. There are two possible cases. First, if track $\tilde{r}_{i+1}$ is empty at instant $t+t_i$, the train enters it immediately and leaves the network at instant $t+t_i+\eta_i$. Second, if track $\tilde{r}_{i+1}$ is not empty, train a_i waits on tracks r_i and $\tilde{r}_i$ until it empties.

With the help of Figure 9.7, one can convince oneself that the behaviour described above corre-

sponds to a Tetris model with matrices :

$$M_{\mathcal{R}}(a_1) = \begin{array}{c} \begin{array}{cccccc} \tilde{r}_1 & r_1 & \tilde{r}_2 & r_2 & \tilde{r}_3 & r_3 \end{array} \\ \begin{pmatrix} t_1 & t_1+\eta_1 & t_1+\eta_1 & \varepsilon & \varepsilon & \varepsilon \\ t_1 & t_1+\eta_1 & t_1+\eta_1 & \varepsilon & \varepsilon & \varepsilon \\ e & \eta_1 & \eta_1 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & e & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & e & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & \varepsilon & \varepsilon & e \end{pmatrix} \end{array} .$$

Matrices $M_{\mathcal{R}}(a_2)$ and $M_{\mathcal{R}}(a_3)$ have the same form, up to the replacement of $\tilde{r}_1, r_1, \tilde{r}_2, t_1, \eta_1$ by $\tilde{r}_i, r_i, \tilde{r}_{i+1}, t_i, \eta_i$.

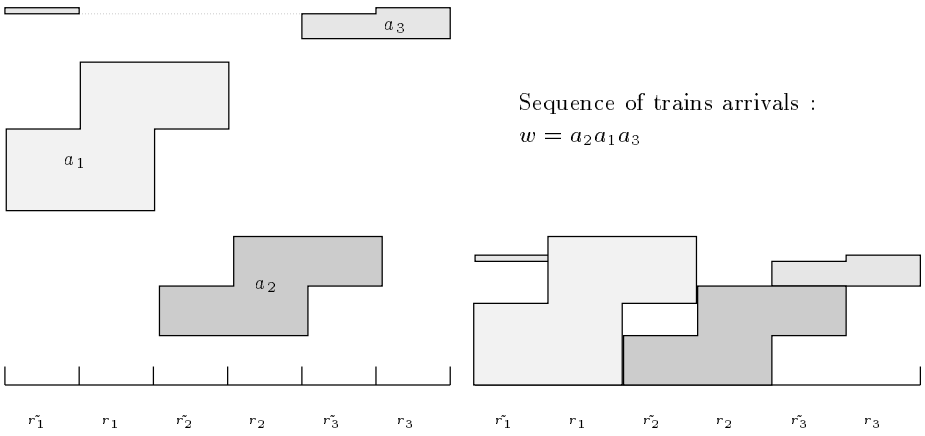


Figure 9.7: Heap of pieces associated with Gagny's triangle.

We propose, in Figure 9.7, the heap of pieces associated with the trains : $w = a_2a_1a_3$. Note that in Figure 9.7 the train of railway 3 is a high speed train (TGV).

9.5 Appendix

Let $\mathbb{M}(\mathcal{A}, D)$ be a trace monoid. We want to find an associated Task Resource model $\mathcal{T} = (\mathcal{A}, \mathcal{R}, R, h \equiv 1)$ with a set $\mathcal{R}$ of minimal cardinality. This will be referred as problem $\mathbb{P}$. We show that problem $\mathbb{P}$ is related to some classical problems in graph theory.

We need to recall some definitions from graph theory. A graph (V, E), $E \subset V \times V$ is *simple* if it contains no loops, i.e. no edge of the form $(v, v), v \in V$. We have already introduced hypergraphs in §9.2.3. In the same way, we say that an hypergraph is simple if it contains no loops. Till the end of the chapter, a graph (resp. hypergraph) is always simple, even if it is not explicitly stated. A graph is complete if $E = (V \times V) - \{\cup(v, v)\}$. The complete graph of order[8] n is denoted by K_n. A clique of a graph is a complete subgraph. A bipartite graph is a graph with two types of vertices and with edges only between vertices of different types. The graph

[8]The order of a graph (or hypergraph) is the number of its vertices.

$K_{p,q}$ is the complete bipartite graph with p vertices of one type and q of the other. Let us define the 2-section of an hypergraph.

Definition 9.5.1. *Let $H = (X, E)$ be an hypergraph. The set of vertices is X and the set of edges is $(E_1, \ldots, E_m)$, $E_i \in \mathcal{P}(X)$. We define its 2-section, denoted by $[H]_2$, as the graph with :*

- *A set of vertices X.*
- *An edge between x_i and x_j iff $i \neq j$, $\exists E_k \in E$ s.t. $x_i, x_j \in E_k$.*

We associate with the trace monoid $\mathbb{M}(\mathcal{A}, D)$ a graph G with vertices $\mathcal{A}$ and edges $\mathcal{D} = \{(a, b) \mid a \neq b, aDb\}$. The problem $\mathbb{P}$ can be reformulated as follows. Given the graph $G = (\mathcal{A}, \mathcal{D})$, find an hypergraph H with a minimal number of edges such that $[H]_2 = G$. The minimal number of edges is denoted by $\Omega(G)$. The interpretation is that each edge corresponds to a resource.

It is immediate that $[G]_2 = G$. It implies that $\Omega(G) \leqslant |\mathcal{D}|$. The equality case is characterized in Theorem 9.5.2. For a proof, see for example Harary [86], p.19 or Berge [20], p.36.

Theorem 9.5.2. *Let $G = (V, E)$ be a connected graph. We have $\Omega(G) = |E|$ if and only if G has no triangles (i.e 3-cliques).*

Next theorem is due to Erdös, Goodman and Pósa [62].

Theorem 9.5.3. *Let $G = (V, E)$ be a connected graph of order n. We have $\Omega(G) \leqslant \left\lfloor \frac{n^2}{4} \right\rfloor$. This bound is the best possible one. It is attained for the graphs $K_{p,p}$ if $n = 2p$ or $K_{p,p+1}$ if $n = 2p + 1$.*

Except for the cases described in Theorem 9.5.2, solving problem $\mathbb{P}$ is NP difficult. Let us give some hints on why it is so. For more details and related results, the reader is referred to Gondran and Minoux [78], ch.10.

1. Let $\mathcal{C} = \{c_i\}$ be a set of cliques of G covering all the edges $\mathcal{D}$. We consider the hypergraph $H = (\mathcal{A}, \mathcal{C})$. We verify that $[H]_2 = G$. Problem $\mathbb{P}$ is equivalent to the search of a minimal covering of G into cliques.

2. We denote by $\mathcal{C} = \{c_i\}$ the set of the *maximal* cliques of G. We consider the hypergraph $H(G) = (\mathcal{A}, \mathcal{C})$.

   We define the matrix B of dimension $|\mathcal{D}| \times |\mathcal{C}|$ by : $B_{dc} = 1$ if $d \in c$ and $B_{dc} = 0$ otherwise. Problem $\mathbb{P}$ reduces to a problem of linear programming in integer numbers :

   $$\begin{cases} \min & \sum_{i=1}^{q} x_i \\ & Bx \geqslant 1 \\ & x \in \{0, 1\}^q \end{cases},$$

   where $q = |\mathcal{C}|$.

3. We consider the associated linear program obtained by replacing $x_i = \{0, 1\}$ by $0 \leqslant x_i \leqslant 1$. There exists a polynomial algorithm to solve it. But, in general, the solutions will not be integer ones. The main case where the algorithm provides integer solutions is when B

is totally unimodular[9], see Hoffman & Kruskal [89]. In general, incidence matrices of hypergraphs are not totally unimodular, hence problem $\mathbb{P}$ can not be solved in polynomial time.

Remark 9.5.4. Incidence matrices of graphs are totally unimodular. However it is not an interesting case. In fact, assume that $H(G)$ is a graph. It implies that the maximal cliques of G are of cardinal strictly less than 3. It means precisely that there are no triangles in G. But we know by Theorem 9.5.2 that problem $\mathbb{P}$ is trivial for such graphs.

We propose in Figure 9.8 an example of a graph G where $\Omega(G)$ is strictly less than the number of maximal cliques of G.

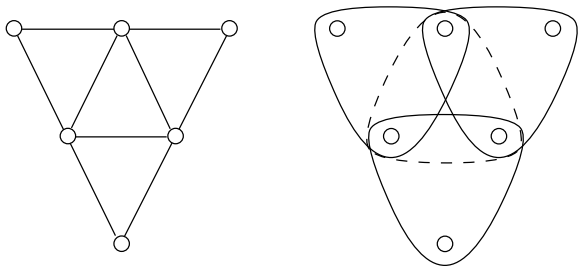


Figure 9.8: Graph with 4 maximal cliques and $\Omega(G) = 3$.

There are four maximal cliques, of order 3. However, the clique represented in doted lines is redundant and we have $\Omega(G) = 3$.

[9] A matrix is totally unimodular if the determinant of each extracted matrix is equal to -1, 0 or 1.

Partie III

Systèmes Non-Linéaires Stochastiques

Chapitre 10

Ergodic Theory of Stochastic Operators and Discrete Event Networks

Théorie Ergodique des Opérateurs et des Réseaux à Événements Discrets Stochastiques

On considère une classe de systèmes généralisant strictement les systèmes (max,+) linéaires. Ces systèmes peuvent être décrits de façon très imprécise comme des opérateurs (min,max,+,×) linéaires. On montre comment les résultats de [4] ainsi que ceux du chapitre 6 peuvent être (partiellement) étendus à cette classe de systèmes. On étudie également les réseaux à événements discrets. On montre que les théorèmes ergodiques démontrés dans le cadre des réseaux à événements discrets peuvent être vus comme une généralisation de ceux obtenus pour les opérateurs. Nous donnons pour conclure divers exemples de systèmes analysables par ce type d'outils.

Ce chapitre provient d'un article réalisé en commun avec François Baccelli, [15]. Les auteurs tiennent à souligner l'importance de l'atelier HP-BRIMS *Idempotency* pour la maturation de ce travail. Cet atelier, organisé par Jeremy Gunawardena, s'est déroulé à Hewlett-Packard Laboratories, Bristol, en octobre 94. Nous tenons également à remercier Serguei Foss et Jean-Marc Vincent pour de nombreuses et instructives discussions sur le sujet.

We present a survey of the main ergodic theory techniques which are used in the study of iterates of monotone and homogeneous stochastic operators. It is shown that ergodic theorems on discrete event networks (queueing networks and/or Petri nets) are a generalization of these stochastic operator theorems. Kingman's subadditive ergodic Theorem is the key tool for deriving what we call first order ergodic results. We also show how to use backward constructions (also called Loynes schemes in network theory) in order to obtain second order ergodic results. We will propose a review of systems entering the framework insisting on two models, precedence constraints networks and Jackson type networks.

10.1 Introduction

Many systems appearing in manufacturing, communication or computer science accept a description in terms of discrete event systems. A usual characteristic of these systems is the existence of some sources of randomness affecting their behaviour. Hence a natural framework to study them is the one of stochastic discrete event systems.
In this paper, we are concerned with two different types of models. First, we consider the study of the iterates $T_n \circ T_{n-1} \circ \cdots \circ T_0$, where $T_i : \mathbb{R}^k \times \Omega \to \mathbb{R}^k$ is a random monotone and homogeneous operator. Second, we introduce and study stochastic discrete event networks entering the so-called monotone-separable framework. A subclass of interest is that of stochastic open discrete event networks.
It will appear that these models, although they have been studied quite independently in the past years, have a lot of common points. They share the same kind of assumptions and properties : monotonicity, homogeneity and non-expansiveness. In fact, we are going to show that monotone-separable discrete event networks are a generalization of monotone-homogeneous operators. However, when a system can be modelled as an operator, it provides a more precise description and stronger results.
In both types of models, we are working with daters. Typically, we have to study a random process $X(n) \in \mathbb{R}^k$, where $X(n)_i$ represents the $n-th$ occurrence of some event in the system. We are going to propose two types of asymptotic results :

1. First order results, concerning the asymptotic rates $\lim_n X(n)_i/n$.

2. Second order results, concerning the asymptotic behaviour of differences such as $X(n)_i - X(n)_j$.

The main references for the results proposed in the paper are the following ones. First order results for operators appear in Vincent [136]. Second order results for operators are new. First and second order results for open discrete event networks are proved in Baccelli and Foss [10]. First order results for general discrete event networks are new. A more complete presentation will be done in a forthcoming paper [**?**].
The paper is organized as follows. In Part I, we treat first order results and in Part II, second order ones. In each part, we consider operators and discrete event networks separately. In a last part, we propose a review of systems entering the frameworks insisting on two models, precedence constraints networks and Jackson type networks.
We aim at emphasizing how theorems on stochastic systems are obtained as an interaction between structural properties of deterministic systems and probabilistic tools. In order to do so,

we introduce first the probabilistic tools (§10.2 and 10.6). Then we present some properties on deterministic systems. At last, we prove the main theorem for stochastic systems.

First Order Ergodic Results

10.2 Probabilistic Tools

We consider a probability space $(\Omega, \mathcal{F}, P)$. We consider a bijective and bi-measurable shift operator $\theta : \Omega \to \Omega$. We assume that θ is stationary and ergodic with respect to the probability P.

Lemma 10.2.1 (Ergodic lemma). *If $\mathcal{A} \in \mathcal{F}$ is such that $\theta(\mathcal{A}) \subset \mathcal{A}$ then $P\{\mathcal{A}\} = 0$ or 1.*

Theorem 10.2.2 (Kingman's subadditive ergodic theorem [97]). *Let $X_{l,n}, l < n \in \mathbb{Z}$, be a doubly-indexed sequence of integrable random variables such that*

- **stationarity** *: $X_{n,n+p} = X_{0,p} \circ \theta^n, \ \forall n, p, \ p > 0$.*
- **boundedness** *: $E[X_{0,n}] \geqslant -Cn, \quad \forall n > 0$, for some finite constant $C > 0$.*
- **subadditivity** *: $X_{l,n} \leqslant X_{l,m} + X_{m,n}, \quad \forall l < m < n$.*

Then there exists a constant γ such that the following convergence holds both in expectation and a.s.

$$\lim_{n\to\infty} \frac{E[X_{0,n}]}{n} = \gamma, \quad \lim_{n\to\infty} \frac{X_{0,n}}{n} = \gamma \quad P \text{ a.s.} \tag{10.1}$$

Remark 10.2.3. The convergence in expectation is straightforward. In fact, we have by subadditivity, $E(X_{0,n}) \leqslant E(X_{0,m}) + E(X_{m,n})$. By stationarity, it implies $E(X_{0,n}) \leqslant E(X_{0,m}) + E(X_{0,n-m})$. The real valued sequence $u_n = \{E(X_{0,n})\}$ is subadditive, hence u_n/n converges in $\mathbb{R} \cup \{-\infty\}$. Because of the boundedness assumption, we conclude that the limit is finite.

Remark 10.2.4. If we have additivity instead of subadditivity, then the previous theorem reduces to the following result:

$$\lim_{n\to\infty} \frac{\Sigma_{i=0}^{n} X_{i,i+1}}{n} \overset{n\to\infty}{\longrightarrow} E(X_{0,1}) \ \ P \text{ a.s.}$$

When the sequence $\{X_{n,n+1}, n \in \mathbb{N}\}$ is i.i.d., this is simply the Strong Law of Large Numbers. More generally, when the sequence $\{X_{n,n+1}, n \in \mathbb{N}\}$ is stationary ergodic (i.e. $X_{n,n+1} = X_{0,1} \circ \theta^n$), it is Birkhoff's ergodic theorem.

10.3 Application to Stochastic Operators

10.3.1 Subadditivity

We call (deterministic) operator a map $T : \mathbb{R}^k \to \mathbb{R}^k$ which is measurable with respect to $\mathcal{B}$, the Borel σ-field of $\mathbb{R}^k$. Let $\{T_n, n \in \mathbb{N}\}$ be a sequence of operators. We associate with it and an initial condition $x_0 \in \mathbb{R}^k$, a sequence on $\mathbb{R}^k$:

$$\left\{ \begin{array}{c} x(n+1) = T_n(x(n)) = T_n \circ \cdots \circ T_0(x(0)) \\ x(0) = x_0 \, . \end{array} \right. \tag{10.2}$$

We will sometimes use the notation $x(n, x_0)$ to emphasize the value of the initial condition.

We consider a probability space $(\Omega, \mathcal{F}, P, \theta)$ as defined above. We call random (or stochastic) operator a map $T : \mathbb{R}^k \times \Omega \to \mathbb{R}^k$ which is measurable with respect to $\mathcal{B} \times \mathcal{F}$. As usual, we will often write $T(x)$ for $T(x, \omega), x \in \mathbb{R}^k, \omega \in \Omega$. A stationary and ergodic sequence of random operators is a sequence $\{T_n, n \in \mathbb{N}\}$ verifying $T_n(x, \omega) = T_0(x, \theta^n \omega)$. In the same way as in Equation (10.2), we associate with $\{T_n, n \in \mathbb{N}\}$ and a (possibly random) initial condition x_0, a random process $\{x(n), n \in \mathbb{N}\}$ taking its values in $\mathbb{R}^k$.

In what follows, definitions apply to deterministic *and* random operators. For random operators, the properties have to be verified with probability 1.

Definition 10.3.1.

1. **Homogeneity** *T is homogeneous if for all $x \in \mathbb{R}^k$ and λ in $\mathbb{R}$, $T(x + \lambda \vec{1}) = \lambda \vec{1} + T(x)$, where $\vec{1}$ is the vector of $\mathbb{R}^k$ with all its coordinates equal to 1.*

2. **Monotonicity** *T is monotone if $x \leqslant y$ implies $T(x) \leqslant T(y)$ coordinatewise.*

For a "physical" interpretation of these conditions, see Remark 10.3.11. The next theorem is a key tool in understanding the importance of homogeneity and monotonicity in what follows.

Theorem 10.3.2 (Crandall-Tartar [48]). *We consider an operator $T : \mathbb{R}^k \to \mathbb{R}^k$ and the following properties*

- H. *T is homogeneous.*
- M. *T is monotone.*
- NE. *T is non-expansive with respect to the sup-norm, i.e $\forall x, y \in \mathbb{R}^k$, $||T(x) - T(y)||_\infty \leqslant ||x - y||_\infty$.*

If H *holds, then there is equivalence between* M *and* NE. *Such operators will be referred to as monotone-homogeneous operators.*

Corollary 10.3.3. *Let us consider a sequence $T_n : \mathbb{R}^k \to \mathbb{R}^k$, $n \in \mathbb{N}$, of monotone-homogeneous operators. If $\exists x \in \mathbb{R}^k$, $\exists i \in \{1, \ldots, k\}$ such that $\lim_n T_n \circ \cdots \circ T_0(x)_i / n$ exists then :*

$$\forall y \in \mathbb{R}^k, \ \lim_n \frac{T_n \circ \cdots \circ T_0(y)_i}{n} = \lim_n \frac{T_n \circ \cdots \circ T_0(x)_i}{n} \, . \tag{10.3}$$

Proof. Straightforward from non-expansiveness

$$\lim_n \frac{||T_n \circ \cdots \circ T_0(y) - T_n \circ \cdots \circ T_0(x)||_\infty}{n} \leqslant \lim_n \frac{||x-y||_\infty}{n}\,.$$

□

Proposition 10.3.4. *Let $T_n : \mathbb{R}^k \to \mathbb{R}^k$ be a sequence of monotone-homogeneous operators. We define $e = (0,\ldots,0)'$ and for $l < n$, $x_{l,n} = T_{n-1} \circ \cdots \circ T_l(e)$. The maximal (resp. minimal) coordinate of $x_{l,n}$ forms a subadditive (resp. super-additive) process, i.e.*

$$\begin{aligned} \forall l < m < n \in \mathbb{N}, \quad \max_i(x_{l,n})_i &\leqslant \max_i(x_{l,m})_i + \max_i(x_{m,n})_i \\ \min_i(x_{l,n})_i &\geqslant \min_i(x_{l,m})_i + \min_i(x_{m,n})_i\,. \end{aligned} \tag{10.4}$$

Proof. We have $\forall l < m < n \in \mathbb{N}$,

$$\begin{aligned} x_{l,n} &= T_{n-1} \circ \cdots \circ T_m \circ T_{m-1} \circ \cdots \circ T_l(e) = T_{n-1} \circ \cdots \circ T_m\,(x_{l,m}) \\ &\leqslant T_{n-1} \circ \cdots \circ T_m \left(e + (\max_i(x_{l,m})_i)\,\vec{1} \right) \quad \text{(monotonicity)} \\ &\leqslant T_{n-1} \circ \cdots \circ T_m(e) + (\max_i(x_{l,m})_i)\,\vec{1} \quad \text{(homogeneity)}. \end{aligned}$$

Therefore,

$$\max_i(x_{l,n})_i \leqslant \max_i(x_{l,m})_i + \max_i(x_{m,n})_i.$$

The proof of the super-additivity of the minimal coordinate is equivalent. □

We are now ready to prove the following theorem on stochastic operators.

Theorem 10.3.5 (Vincent [136]). *Let $\{T_n, n \in \mathbb{N}\}$ be a stationary ergodic sequence of monotone-homogeneous random operators. We define the process $x(n,y), y \in \mathbb{R}^k$, as in Equation (10.2). If, for all n, the random variable $T_n \circ \cdots \circ T_1(0)$ is integrable and such that $E(T_n \circ \cdots \circ T_1(0)) > -Cn$, for some positive C, then $\exists \overline{\gamma}, \underline{\gamma} \in \mathbb{R}$ such that $\forall y \in \mathbb{R}^k$,*

$$\lim_n \frac{\max_i x(n,y)_i}{n} = \overline{\gamma} \;\; P \text{ a.s.}, \quad \lim_n \frac{E(\max_i x(n,y)_i)}{n} = \overline{\gamma} \tag{10.5}$$

$$\lim_n \frac{\min_i x(n,y)_i}{n} = \underline{\gamma} \;\; P \text{ a.s.}, \quad \lim_n \frac{E(\min_i x(n,y)_i)}{n} = \underline{\gamma} \tag{10.6}$$

Proof. We define as previously the doubly-indexed sequence $x_{l,n} = T_{n-1} \circ \cdots \circ T_l(e)_i$, $l < n$. Using Prop. 10.3.4, the sequences $\max_i(x_{n,m})_i$ and $-\min_i(x_{n,m})_i$ are subadditive. Hence they satisfy the conditions of Theorem 10.2.2. So Equation (10.5) holds for $y = e = (0,\ldots,0)'$. For any other initial condition y, we obtain $\lim_n x(n,y)/n = \lim_n x(n,e)/n$ using the non-expansiveness as in Corollary 10.3.3. □

The convergence for the maximal and minimal rates does not imply that of the coordinates. Here is a counter-example borrowed from [136].

Example 10.3.6. We consider a random operator $T_0 : \mathbb{R}^3 \to \mathbb{R}^3$ verifying:

$$x = (x_1, x_2, x_3)',\ T_0(x) = (x_1 + 1, x_2 + 2, U_0 x_1 + (1 - U_0)x_2)' ,$$

where U_0 is a $[0,1]$-uniform random variable. We have

$$\liminf (T_n \dots T_0(x)_3)/n = 1 \quad \text{and} \quad \limsup (T_n \dots T_0(x)_3)/n = 2 .$$

Here is another example of the same kind:

$$\begin{aligned} T_0(x) \quad = \quad & (\delta_0(\max(x_1, x_2) + 2) + (1 - \delta_0)(\min(x_1, x_2) + 1), (1 - \delta_0) \\ & (\max(x_1, x_2) + 2) + \delta_0(\min(x_1, x_2) + 1), U_0 x_1 + (1 - U_0)x_2)' , \end{aligned}$$

where U_0 is a $[0,1]$-uniform random variable and δ_0 is a (0,1) Bernouilli random variable. The random variables U_0 and δ_0 are independent.

10.3.2 Projective boundedness

In order to complete Proposition 10.3.4 or Theorem 10.3.5, the two main questions are :

i. Does a limit exist for the vector $(T_n \circ \dots \circ T_0(y)_1/n, \dots, T_n \circ \dots \circ T_0(y)_k/n)\ \Gamma$

ii. Is this limit equal to a "constant" $(\gamma, \dots, \gamma)\ \Gamma$

The general answers to these questions are not known (even for deterministic operators). We are going to propose a sufficient condition to answer positively *i.* and *ii.* Let us introduce some definitions.

Definition 10.3.7 ($\mathbb{PR}^k$). *We consider the parallelism relation :*

$$u, v \in \mathbb{R}^k \quad u \simeq v \iff \exists a \in \mathbb{R} \text{ such that } \forall i,\ u_i = a + v_i .$$

We define the projective space $\mathbb{PR}^k$ as the quotient of $\mathbb{R}^k$ by this parallelism relation. Let π be the canonical projection of $\mathbb{R}^k$ into $\mathbb{PR}^k$.

Definition 10.3.8. *Let T be an operator of $\mathbb{R}^k$ into $\mathbb{R}^k$.*

1. *T is projectively bounded if $\exists K$ a compact of $\mathbb{PR}^k$ such that the image of T is included in K, i.e. $\pi(\mathrm{Im}(T)) \subset K$.*
2. *T has a generalized fixed point if $\exists \gamma \in \mathbb{R}, x_0 \in \mathbb{R}^k$ such that $T(x_0) = \gamma \vec{1} + x_0$. It is equivalent to say that T has a fixed point in the projective space (see Def. 10.3.7).*

Proposition 10.3.9. *Let us consider $T : \mathbb{R}^k \to \mathbb{R}^k$ a monotone-homogeneous operator. Let us consider the following assumptions.*

A. T is projectively bounded.
B. T has a generalized fixed point.
C. $\forall x$, $\lim_n T^n(x)/n = (\gamma, \dots, \gamma)'$.

The following implications hold : $A \Rightarrow B \Rightarrow C$. The other implications are false, $C \not\Rightarrow B \not\Rightarrow A$ and $C \not\Rightarrow A$.

Proof. 1. $A \Rightarrow B$. Let K be a compact of $\mathbb{PR}^k$ such that $\pi(T(\mathbb{R}^k)) \subset K$. It implies that $\pi(T) : K \to K$. Hence $\pi(T)$ is continuous on a compact and has a fixed point by application of Brouwer's Theorem.

2. $B \Rightarrow C$. Let $x \in \mathbb{R}^k$ be a generalized fixed point of T, i.e $T(x) = \gamma\vec{1} + x$. It implies $T^n(x) = n\gamma\vec{1} + x$ and $\lim_n T^n(x)/n = (\gamma, \ldots, \gamma)'$. From Corollary 10.3.3, we have $\forall y \in \mathbb{R}^k, \lim_n T^n(y)/n = (\gamma, \ldots, \gamma)'$.

3. $B \not\Rightarrow A$ and $C \not\Rightarrow A$. An easy counter-example is obtained by considering the identity operator $I : \mathbb{R}^k \to \mathbb{R}^k$, $I(x) = x$.

4. $C \not\Rightarrow B$ There exist counter-examples of dimension 2, [83].

□

This Proposition has an interesting application for stochastic operators.

Theorem 10.3.10. *Let $\{T_n, n \in \mathbb{N}\}$ be a stationary and ergodic sequence of random operators. We assume that there exist $l \in \mathbb{N}$ and K a compact of $\mathbb{PR}^k$ such that :*

$$\mathcal{E} = \{\pi(\mathrm{Im}(T_{l-1} \circ \cdots \circ T_0)) \subset K\}\,, \quad P(\mathcal{E}) > 0\,. \tag{10.7}$$

Then $\exists \gamma \in \mathbb{R}$, such that

$$\forall x \in \mathbb{R}^k, \quad \lim_n \frac{T_n \circ \cdots \circ T_0(x)}{n} = (\gamma, \ldots, \gamma)'\ .$$

Proof. We define recursively the random variables

$$\begin{aligned} N_1 &= \min\{n \in \mathbb{N} \,|\, T_{n+l-1} \circ \cdots \circ T_n \in \mathcal{E}\}, \\ N_{i+1} &= \min\{n \in \mathbb{N} \,|\, n \geqslant N_i + l,\ T_{n+l-1} \circ \cdots \circ T_n \in \mathcal{E}\}\,. \end{aligned}$$

First of all, let us prove that the random variables N_i are almost surely finite. Let us consider the event $\mathcal{A}_1 = \{N_1 < +\infty\}$. It is easy to see that $\mathcal{A}_1$ is invariant by the shift θ. In fact $N_1(\theta^{-1}\omega) = N_1(\omega) + 1$ or 0. Hence $\{N_1(\omega) < +\infty\} \Rightarrow \{N_1(\theta^{-1}\omega) < +\infty\}$, i.e. $\theta(\mathcal{A}) \subset \mathcal{A}$. By Lemma 10.2.1, it implies that $\mathcal{A}$ is of probability 0 or 1. But $(\{N_1 = 0\} = \mathcal{E}) \subset \mathcal{A}$ and by assumption $P(\mathcal{E}) > 0$. We conclude that $P(\mathcal{A}) = 1$. A similar argument can now be applied to N_2. For $\mathcal{A} \in \mathcal{F}$, we define the indicator function $\mathbf{1}_{\mathcal{A}} : \Omega \to \Omega$, $\mathbf{1}_{\mathcal{A}}(\omega) = 1$ iff $\omega \in \mathcal{A}$. We have

$$\begin{aligned} P(N_2 < +\infty) &= E(\mathbf{1}_{\{N_2<+\infty\}}) = E(\sum_k \mathbf{1}_{\{N_1=k\}}\mathbf{1}_{\{N_2<+\infty\}}) \\ &= E(\sum_k \mathbf{1}_{\{N_1=k\}}\mathbf{1}_{\{N_1\circ\theta^{k+l}<+\infty\}}) = E(\sum_k \mathbf{1}_{\{N_1=k\}}) = 1\,. \end{aligned}$$

We conclude the proof by induction.
Let $\overline{\gamma}$ and $\underline{\gamma}$ be the maximal and minimal rates as defined in Theorem 10.3.5. Let us assume that $\overline{\gamma} \neq \underline{\gamma}$. It implies, $\forall x \in \mathbb{R}^k$,

$$\liminf_n (\max_i x(n)_i - \min_i x(n)_i) = +\infty\,. \tag{10.8}$$

But we also have that $\forall i \in \mathbb{N}, \pi(x(N_i+l)) \subset K$. It implies that $(\max_j x(N_i+l)_j - \min_j x(N_i+l)_j) \subset K'$ where K' is a compact of $\mathbb{R}$. Hence there exists a subsequence $N_{\sigma(i)}$ such that $(\max_j x(N_{\sigma(i)}+l)_j - \min_j x(N_{\sigma(i)}+l)_j)$ converges to a finite limit. This is in contradiction with (10.8). □

Remark 10.3.11. In many applications, the operator will be applied on a vector of dates for a physical system. The vectors $x(n)$ and $x(n+1) = T_n(x(n))$ will represent the dates of the n-th and $(n+1)$-th occurrences of some events in a system. In such a case, the homogeneity property can be interpreted as the fact that changing the absolute origin of times does not modify the dynamic of the system. Hence it becomes a very natural assumption. The monotonicity is interpreted as the fact that delaying an event delays all following events.

10.4 Application to Stochastic Discrete Event Networks

10.4.1 Discrete event networks

A *discrete event network* is characterized by

1. A sequence
$$N = N_{[-\infty,\infty]} = \{\sigma(k), M(k),\ k \in \mathbb{Z}\},$$
where $\sigma(k) \in \mathbb{R}^+$ and $\{M(k)\}$ is a sequence of F-valued variables, where F is some measurable space. With N and $n \leqslant m \in \mathbb{Z}$, we associate the sequence $N_{[n,m]}$ defined by:
$$N_{[n,m]} \stackrel{\text{def}}{=} \{\sigma_{[n,m]}(k), M(n+k),\ k \in \mathbb{N}\},$$
where $\sigma_{[n,m]}(k) \stackrel{\text{def}}{=} \sigma(n+k)$, for $0 \leqslant k \leqslant m-n$, and $\sigma_{[n,m]}(k) \stackrel{\text{def}}{=} \infty$, for $k > m-n$.

2. Measurable functions $\Phi(k,.)$ and $\Psi(.)$: $(\mathbb{R}^+ \times F)^{\mathbb{N}} \to \mathbb{R} \cup \{\infty\}$, $k \in \mathbb{N}^*$, through which are defined
$$X_{[n,m]} = \Psi(N_{[n,m]}), \quad n \leqslant m, \quad X^-_{[n,m]}(k) = \Phi(k, N_{[n,m]}), \quad k \geqslant 1.$$

Remark These variables receive the following interpretations: $X^-_{[n,m]}(k)$ is the initiation date of the k-th event on some reference node, for the *driving sequence* $N_{[n,m]}$.

$$X^+_{[n,m]}(k) \stackrel{\text{def}}{=} X^-_{[n,m]}(k) + \sigma_{[n,m]}(k), \quad n \leqslant m,\ k \geqslant 0$$

is the completion date of this event. $X^+_{[n,m]}(k)$ and $X^-_{[n,m]}(k)$ are called *internal daters*. $X_{[n,m]}$ is the *maximal dater*, i.e. the date of the last event in the network, for the sequence $N_{[n,m]}$.

10.4.2 The monotone–separable framework

Let N and $\widetilde{N}$ be two driving sequences such that $\sigma(k) \leqslant \widetilde{\sigma}(k) < \infty$ for all k, and with $M(k) = \widetilde{M}(k)$ for all k. We denote $X^-_{[1,m]}(k)$, $X^+_{[1,m]}(k)$ and $X_{[1,m]}$ the daters associated with $N_{[1,m]}$ and $\widetilde{X}^-_{[1,m]}(k)$, etc. those associated with $\widetilde{N}_{[1,m]}$.
A network is said to be *monotone-separable* if it satisfies the following properties for all $m \geqslant 1$, $k \geqslant 1$ and for all N and $\widetilde{N}$ as above:

- **causality** $X^-_{[1,m]}(m+1) \leqslant X_{[1,m]} < \infty$.
- **monotonicity** $X^-_{[1,m]}(k) \leqslant \widetilde{X}^-_{[1,m]}(k)$ and $X_{[1,m]} \leqslant \widetilde{X}_{[1,m]}$.
- **non-expansiveness** [1] $\widetilde{X}^-_{[1,m]}(k) - X^-_{[1,m]}(k) \leqslant x$ and $\widetilde{X}_{[1,m]} - X_{[1,m]} \leqslant x$, if $\widetilde{\sigma}(k) = \sigma(k)$ for all $k \neq l$, and $\widetilde{\sigma}(l) = \sigma(l) + x$, $x > 0$.
- **separability** For $1 \leqslant l < m$, if $X_{[1,l]} \leqslant X^+_{[1,m]}(l+1)$ then $X_{[1,m]} \leqslant X^-_{[1,m]}(l+1) + X_{[l+1,m]}$.

Proposition 10.4.1. *Under the above assumptions, the sequence $X_{[m,n]}$ satisfies the sub-additive inequality*

$$X_{[m,n]} \leqslant X_{[m,l]} + X_{[l+1,n]}, \quad \forall m \leqslant l < n.$$

Proof. It is enough to prove the property for $m = 1$, since the general relation will then be obtained by applying the relation for $m = 1$ to the variables associated with some adequate sequence. Let $1 \leqslant l < n$. There are two cases:

Case 1: $X_{[1,l]} \leqslant X^+_{[1,n]}(l+1)$. Then, in view of separability

$$\begin{aligned} X_{[1,n]} &\leqslant X^-_{[1,n]}(l+1) + X_{[l+1,n]} \\ &\leqslant X_{[1,l]} + X_{[l+1,n]}, \end{aligned}$$

where we used the fact that $X_{[1,l]} \geqslant X^-_{[1,l]}(l+1) \geqslant X^-_{[1,n]}(l+1)$, which follows from causality and monotonicity ($X^-_{[1,l]}(l+1) = \widetilde{X}^-_{[1,n]}(l+1)$ with $\widetilde{\sigma}(k) = \sigma(k)$ for $1 \leqslant k \leqslant l$ and $\widetilde{\sigma}(k) = \infty$ for $k > l$).
Case 2: $X_{[1,l]} > X^+_{[1,n]}(l+1)$.
Consider the two sequences $\{\sigma(k)\}$ and $\{\widetilde{\sigma}(k)\}$, which only differ in their $(l+1)$-st coordinate, for which we take $\widetilde{\sigma}(l+1) = \sigma(l+1) + x$, $x > 0$. In view of monotonicity, $X_{[1,n]} \leqslant \widetilde{X}_{[1,n]}$. In

[1] If one sees $(\Psi(.), \Phi(k,.),\ k \geqslant 1)$ as a function: $(\mathbb{R}^+)^{\mathbb{N}} \to (\mathbb{R} \cup \{\infty\})^{\mathbb{N}}$ – the sequence $\{M(k)\}$ being fixed – this is indeed non-expansiveness when taking a L^1 norm on $(\mathbb{R}^+)^{\mathbb{N}}$ and a L^∞ norm on $(\mathbb{R} \cup \{\infty\})^{\mathbb{N}}$.

particular, if we take $x = x^*$ with $x^* = X_{[1,l]} - X^+_{[1,n]}(l+1) > 0$, then

$$\begin{aligned}
\widetilde{X}^+_{[1,n]}(l+1) &= \widetilde{X}^-_{[1,n]}(l+1) + \sigma(l+1) + x^* \\
&= \widetilde{X}^-_{[1,n]}(l+1) + \sigma(l+1) + X_{[1,l]} - X^+_{[1,n]}(l+1) \\
&= X_{[1,l]} + \widetilde{X}^-_{[1,n]}(l+1) - X^-_{[1,n]}(l+1). \qquad (10.9)
\end{aligned}$$

But $X_{[1,l]}$ does not depend on $\sigma(l+1)$, and so $X_{[1,l]} = \widetilde{X}_{[1,l]}$. Therefore

$$\begin{aligned}
\widetilde{X}^+_{[1,n]}(l+1) &= \widetilde{X}_{[1,l]} + \widetilde{X}^-_{[1,n]}(l+1) - X^-_{[1,n]}(l+1) \\
&\geqslant \widetilde{X}_{[1,l]} \quad \text{(monotonicity)}
\end{aligned}$$

We finally obtain that, for $x = x^*$

$$\begin{aligned}
\widetilde{X}_{[1,n]} &\leqslant \widetilde{X}^-_{[1,n]}(l+1) + \widetilde{X}_{[l+1,n]}, \quad \text{(separability)} \\
&= \widetilde{X}^+_{[1,n]}(l+1) + X^-_{[1,n]}(l+1) - X_{[1,l]} + \widetilde{X}_{[l+1,n]}, \quad \text{(Equation (10.9))} \\
&\leqslant \widetilde{X}^+_{[1,n]}(l+1) + X^-_{[1,n]}(l+1) - X_{[1,l]} + x^* + X_{[l+1,n]}, \quad (\text{non} - \text{exp.}) \\
&= \widetilde{X}^+_{[1,n]}(l+1) + X^-_{[1,n]}(l+1) - X_{[1,l]} + X_{[1,l]} - X^+_{[1,n]}(l+1) + X_{[l+1,n]} \\
&= \widetilde{X}^+_{[1,n]}(l+1) - X^+_{[1,n]}(l+1) + X^-_{[1,n]}(l+1) + X_{[l+1,n]} \\
&\leqslant x^* + X^-_{[1,n]}(l+1) + X_{[l+1,n]}, \quad (\text{non} - \text{exp.}) \\
&= X^-_{[1,n]}(l+1) - X^+_{[1,n]}(l+1) + X_{[1,l]} + X_{[l+1,n]} \\
&\leqslant X_{[1,l]} + X_{[l+1,n]}.
\end{aligned}$$

□

Remark 10.4.2. Under the additional assumption that $X^-_{[1,m]}(l+1)$ is a function of $\{\sigma(k), 1 \leqslant k \leqslant l$, and $M(p)$, $1 \leqslant p \leqslant m\}$ only, non-expansiveness can be replaced by the following property:

- **sub-homogeneity** $\widetilde{X}_{[1,m]} \leqslant X_{[1,m]} + \lambda$, if $\widetilde{\sigma}(1) = \sigma(1) + \lambda$ and $\widetilde{\sigma}(k) = \sigma(k)$ for all $k > 1$, $\lambda > 0$ and $m \geqslant 1$.

The proof is exactly the same for case 1. For case 2, taking x^* as in the proof of Proposition 10.4.1 gives $\widetilde{X}^+_{[1,n]}(l+1) = X_{[1,l]}$ and

$$\begin{aligned}
\widetilde{X}_{[1,n]} &\leqslant \widetilde{X}^-_{[1,n]}(l+1) + \widetilde{X}_{[l+1,n]}, \quad \text{(separability)} \\
&= X^-_{[1,n]}(l+1) + \widetilde{X}_{[l+1,n]} \\
&\leqslant X^-_{[1,n]}(l+1) + X_{[l+1,n]} + X_{[1,l]} - X^+_{[1,n]}(l+1), \quad (\text{sub} - \text{homog.}) \\
&\leqslant X_{[1,l]} + X_{[l+1,n]}.
\end{aligned}$$

Remark 10.4.3. Some generalizations of the framework, with internal daters, will be proposed in [**?**]. See also the Jackson network example of §10.9.2. The comments on the physical interpretation of homogeneity or monotonicity made in Remark 10.3.11 also apply to discrete event networks.

10.4.3 Open discrete event networks

A discrete event network is said to be *open* if the following additional assumption holds for all $m \geqslant 1$:

$$\forall 1 \leqslant k \leqslant m,\ X^-_{[1,m]}(k+1) = X^-_{[1,\infty]}(k+1) = X^+_{[1,m]}(k), \text{ and } X^-_{[1,m]}(1) = 0.$$

One can then define a point process $\{A_k\}_{k\geqslant 1}$ by

$$A_k = A_1 + X^-_{[1,\infty]}(k).$$

The origin of this point process is arbitrary. It is then possible to interpret $\{A_k\}$ as an external arrival process, the inter-arrival times being the sequence $\{\sigma(k)\}$. To summarize, an open discrete network is described by a sequence $N = N_{[-\infty,\infty]} = \{A_k, M(k),\ k \in \mathbb{Z}\}$.
The conditions of the monotone separable framework take the following form for an open network (which corresponds to the conditions of [10]) : for all $m \geqslant 1$, the following properties hold:

- **causality** $A_m \leqslant A_1 + X_{[1,m]} < \infty$.
- **monotonicity** $\widetilde{X}_{[1,m]} \geqslant X_{[1,m]}$, for $\widetilde{N}$ and N with $\widetilde{\sigma}(k) \geqslant \sigma(k)$ for all k.
- **homogeneity** Let $\widetilde{N}$ be the point process obtained by shifting the points of N A_k, $k \geqslant 1$, of $\lambda > 0$ to the right. Then $\widetilde{X}_{[1,n]} = X_{[1,n]}$.
- **separability** $A_1 + X_{[1,m]} \leqslant A_{l+1} + X_{[l+1,m]}$ for all $1 \leqslant l < m$ such that $A_1 + X_{[1,l]} \leqslant A_{l+1}$.

For an open network, monotonicity can be interpreted as the fact that delaying an arrival delays all forthcoming events in the network. For a possible interpretation of separability, see Remark 10.7.1.

10.4.4 Stochastic discrete event networks

We consider a probability space $(\Omega, \mathcal{F}, P, \theta)$ as in §10.2. The following stochastic assumptions are made:

- **compatibility** $(\sigma(k), M(k)) = (\sigma(0), M(0)) \circ \theta^k$ for all $k \in \mathbb{Z}$.
- **integrability** $\exists C > 0$, $-Cm \leqslant E[X_{[1,m]}] < \infty$ for all $m \geqslant 0$.

Theorem 10.4.4. *For all discrete event network which satisfies the monotone-separable assumptions and the above stochastic assumptions, we have*

$$\lim_{n\to\infty} \frac{X_{[1,n]}}{n} = \gamma \quad \text{a.s.} \quad \text{and} \lim_{n\to\infty} \frac{E[X_{[1,n]}]}{n} = \gamma \tag{10.10}$$

for some finite constant γ.

Proof. We have $X_{[m,m+p]} = X_{[0,p]} \circ \theta^m$, for all $m \in \mathbb{Z}$ and $p \geqslant 0$. For $m \leqslant n$, define $Y_{[m,n+1]} = X_{[m,n]}$. From Proposition 10.4.1, for all $m \leqslant l < n$,

$$Y_{[m,n+1]} \leqslant Y_{[m,l+1]} + Y_{[l+1,n+1]}.$$

So $\{Y_{[m,n]}\}$, $m < n$, satisfies all the assumptions of Theorem 10.2.2. □

10.5 Relations Between Operators and Networks

Let us investigate the relation between the operator framework considered in §10.3 and the monotone-separable framework considered above. Let $\{T_n\}$ be a sequence of monotone and homogeneous operators. Let $\sigma(n) \equiv 0$ and $M(n) = T_n$. Let $x(n,0)$ be the variables associated with the operator recurrence equation (10.2) with initial condition $x_0 = 0$. With these variables, we associate

$$X^-_{[0,n]}(k) = X^+_{[0,n]}(k) = \max_i x(k-1,0)_i, \quad k \geqslant 1, \quad \text{and} \quad X_{[0,n]} = \max_i x(n,0)_i,$$

Note that these variables are functions of $\{M(l)\}$. We have

- $X^-_{[0,n]}(n+1) = X_{[0,n]} < \infty$, so that causality holds.
- Monotonicity and non-expansiveness trivially hold as neither $X^-_{[0,n]}(k)$, $k \geqslant 1$, nor $X_{[0,n]}$ depend upon $\{\sigma(l)\}$.
- Separability holds because it is always true that $X_{[0,l]} = X^-_{[0,m]}(l+1)$ and

$$\begin{aligned}
X_{[0,m]} &= \max_i \left(T_m \circ \ldots \circ T_{l+1}(x(l,0))\right)_i \\
&= \max_i \left(T_m \circ \ldots \circ T_{l+1}(x(l,0) + (X_{[0,l]} - X_{[0,l]})\vec{1})\right)_i \\
&= X_{[0,l]} + \max_i \left(T_m \circ \ldots \circ T_{l+1}(x(l,0) - X_{[0,l]}\vec{1})\right)_i, \quad \text{(homog.)} \\
&\leqslant X_{[0,l]} + \max_i \left(T_m \circ \ldots \circ T_{l+1}(0)\right)_i, \quad \text{(monotonicity)} \\
&= X^-_{[0,m]}(l+1) + X_{[l+1,m]}.
\end{aligned}$$

Hence, monotone separable operators are a special case of monotone separable discrete event networks. On the other hand, it should be remarked that an operator can *not* be represented as an *open* discrete event network. A representation in terms of operators is interesting as it is more precise than the corresponding discrete event network one. In particular, we will see that we are able to obtain second order results for operators, §10.8, and not for non-open discrete event networks, §10.7.2.

Second Order Ergodic Results

We will introduce a construction which is known as the Loynes scheme. This type of construction will be used for both types of models, discrete event networks and operators, but in a rather different way.

10.6 Basic Example and Probabilistic Tools

The basic construction was introduced by Loynes in [103] to study the stability of the $G/G/1/\infty$ queue. A G/G arrival process is a stationary and ergodic marked point process

$$N = \{(\tau_n, \sigma_n), n \in \mathbb{Z}\},$$

where $\sigma(n) \in \mathbb{R}^+$ is the service time required by customer n and $\tau_n = A_{n+1} - A_n$ the inter-arrival time between customers n and $n+1$. The $1/\infty$ part describes the queueing mechanism. There is a single server and an infinite waiting room or buffer. Upon arrival at instant A_n, customer n is served immediately if the server is idle at A_n^- and is queued in the buffer otherwise. The server operates at unit rate until all customers present in the buffer have been served. Let $X_{[l,n]}$ be the time of last activity in the system, i.e. the departure of the last customer, for the restriction $N_{[l,n]}$. Here are two equivalent ways to describe the system :

- As a stochastic operator,

$$\begin{pmatrix} A_{n+1} \\ X_{[l,n+1]} \end{pmatrix} = \begin{pmatrix} \tau_n + A_n \\ \max(\tau_n + \sigma_{n+1} + A_n \ , \ \sigma_{n+1} + X_{[l,n]}) \end{pmatrix} \tag{10.11}$$

$$= \begin{pmatrix} \tau_n & \varepsilon \\ \tau_n \otimes \sigma_{n+1} & \sigma_{n+1} \end{pmatrix} \otimes \begin{pmatrix} A_n \\ X_{[l,n]} \end{pmatrix} \tag{10.12}$$

Equation (10.11) can be written $X_{[l,n+1]} = \max(A_{n+1}, X_{[l,n]}) + \sigma_{n+1}$. The meaning is that the server starts working on customer $n+1$ as soon as this customer has arrived (A_{n+1}) *and* the server has completed the services of the previous customers ($X_{[l,n]}$). Equation (10.12) is just a re-writing using the (max,+) notations, see also §10.9.1. It is easy to verify that this operator is monotone and homogeneous.

- As an open network, by means of the function Ψ of §10.4.

$$X_{[l,n]} = \Psi(\tau_i, \sigma_i, i \in \{l, \ldots, n\})$$

$$= (A_n - A_l) + \sigma_n + \max(0, \max_{k=1}^{n-l} \sum_{i=1}^{k} (\sigma_{n-i} - \tau_{n-i})) \ . \tag{10.13}$$

The easiest way to understand Equation (10.13) is to look at Figure 10.1. Function Ψ is monotone, homogeneous and separable.
Let us consider the sequence of variables $\{Z_{[l,n]},\ l \leqslant n \in \mathbb{Z}\}$ defined by $Z_{[l,n]} = X_{[l,n]} - (A_n - A_l)$. The variables $Z_{[l,n]}$ verify Lindley's equation[2] $Z_{[l,n+1]} = (Z_{[l,n]} - \tau_n)^+ + \sigma_{n+1}$.

Theorem 10.6.1 (Loynes [103]). *The sequence $Z_{[-n,0]}$ is increasing in n, i.e. $Z_{[-n-1,0]} \geqslant Z_{[-n,0]}$. The limit $Z = \lim_n Z_{[-n,0]}$ verifies $P\{Z < +\infty\} = 0$ or 1. Furthermore Z is a stationary solution of Lindley's equation, i.e. $Z(\theta\omega) = (Z(\omega) - \tau_0)^+ + \sigma_1$. When $P\{Z < +\infty\} = 1$, the sequence $\{Z_{[0,n]}, n \in \mathbb{N}\}$ couples in finite time with the stationary sequence $\{Z \circ \theta^n\}$.*

[2] It is more classical, but equivalent, to work with the workload variable $W_n = X_{[0,n]} - A_n - \sigma(n)$, yielding equation $W_{n+1} = (W_n + \sigma_n - \tau_n)^+$.

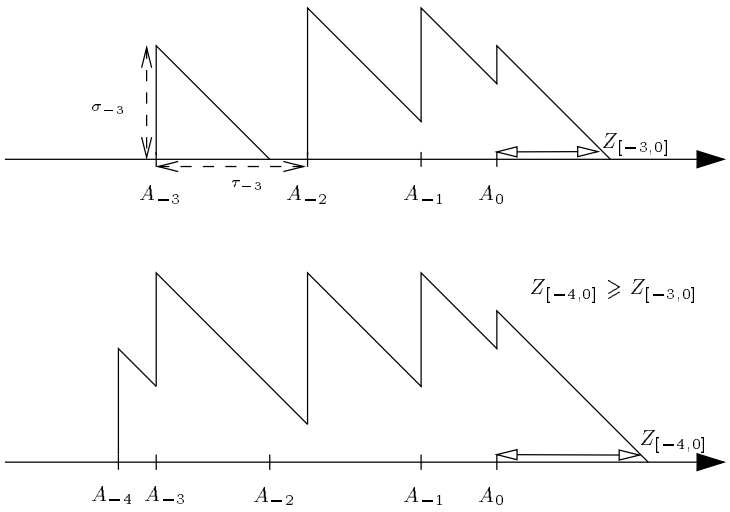


Figure 10.1: Loynes scheme for the $G/G/1/\infty$ queue.

Proof. The monotonicity of $Z_{[-n,0]}$ is easy to obtain from Equation (10.13). It is also illustrated in Figure 10.1. Hence the limit $Z = \lim_n Z_{[-n,0]}$ exists. Let us denote $\mathcal{A} = \{Z < +\infty\}$. From $Z_{[-n,1]} = (Z_{[-n,0]} - \tau_0)^+ + \sigma_1$ and the fact that σ_1 is a.s. finite, we obtain

$$Z(\omega) < +\infty \Leftrightarrow \exists K \ \forall n,\ Z_{[-n,0]}(\omega) < K \Rightarrow \exists K' \ \forall n,\ Z_{[-n,1]}(\omega) < K' .$$

But we also have

$$Z_{[-n,1]}(\omega) = Z_{[-n-1,0]}(\theta\omega) . \tag{10.14}$$

We conclude that $Z(\theta\omega) < +\infty$. We have proved that $\theta(\mathcal{A}) \subset \mathcal{A}$ which implies, Ergodic Lemma 10.2.1, that $P\{\mathcal{A}\} = 0$ or 1. From Equation (10.14), letting n go to ∞, we deduce that $Z(\theta\omega) = (Z(\omega) - \tau_0)^+ + \sigma_1$. For a proof of the remaining point, see [103] or [6]. □

The limit Z is usually referred to as Loynes variable. We can obtain, using Equation (10.13), $P\{Z < +\infty\} = 1 \Leftrightarrow E(\sigma) < E(\tau)$. The condition $E(\sigma) < E(\tau)$ defines the stability region. This condition is usually written under the form $\rho = E(\sigma)/E(\tau) < 1$. We will see a similar type of stability condition in Theorem 10.7.3.

10.7 Application to Stochastic Discrete Event Networks

10.7.1 Open discrete event networks

The assumptions and notations are those of §10.4.3 but we replace the separability assumption by

- **strong separability** For $1 \leqslant l < m$, if $A_1 + X_{[1,l]} \leqslant A_{l+1}$ then $A_1 + X_{[1,m]} = A_{l+1} + X_{[l+1,m]}$.

Remark 10.7.1. Strong separability can be interpreted as follows. If the arrival of customer $l+1$ takes place later than the last activity for the arrival process $[1,l]$, then the evolution of the network after time A_{l+1} is the same as in the network which starts "empty" at this time.

We define $\lambda = E(A_{n+1} - A_n)^{-1}$ interpreted as the arrival rate and

$$Z_{[l,n]} = X_{[l,n]} - (A_n - A_l), \quad l \leqslant n\ . \tag{10.15}$$

Proposition 10.7.2 (Internal monotonicity). *Under the above assumptions, we have*

$$Z_{[l-1,n]} \geqslant Z_{[l,n]}, \quad l \leqslant n.$$

Proof. Consider the point process $\widetilde{N}$ with $\widetilde{\sigma}(l-1) = \sigma(l-1) + Z_{[l-1,l-1]}$ and $\widetilde{\sigma}(k) = \sigma(k)$ everywhere else. For $\widetilde{N}_{[l-1,n]}$, we have separability in l so that

$$\begin{aligned}
\widetilde{X}_{[l-1,n]} &= \widetilde{X}_{[l,n]} + \widetilde{A}_l - \widetilde{A}_{l-1} \\
&= X_{[l,n]} + \widetilde{A}_l - \widetilde{A}_{l-1} \quad (\text{strong} - \text{separability}) \\
&= X_{[l,n]} + A_l - A_{l-1} + Z_{[l-1,l-1]}\ .
\end{aligned} \tag{10.16}$$

Therefore

$$\begin{aligned}
Z_{[l-1,n]} &= X_{[l-1,n]} - (A_n - A_{l-1}) \\
&= X_{[l-1,n]} - (A_n - A_l) - (A_l - A_{l-1}) \\
&= X_{[l-1,n]} - (A_n - A_l) + X_{[l,n]} - \widetilde{X}_{[l-1,n]} + Z_{[l-1,l-1]} \quad (\text{by (10.16)}) \\
&= Z_{[l,n]} + X_{[l-1,n]} - \widetilde{X}_{[l-1,n]} + Z_{[l-1,l-1]} \\
&\geqslant Z_{[l,n]}, \quad (\text{non} - \text{expansiveness}).
\end{aligned}$$

□

Let $Z = \lim_n Z_{[-n,0]}(N)$, which exists by internal monotonicity of $Z_{[-n,0]}(N)$. We define a c-scaling of the arrival point process N in the following way :

$$0 \leqslant c < +\infty, \quad cN = \{cA_n, M(n), n \in \mathbb{Z}\}\ .$$

From Equation (10.15) and Prop. 10.4.1 , we obtain that $Z_{[1,n]}$ is subadditive. Applying Theorem 10.4.4, we obtain the existence of the limits

$$\lim_n \frac{Z_{[1,n]}(cN)}{n} = \lim_n \frac{Z_{[-n,0]}(cN)}{n} = \gamma(c)\ .$$

From Equation (10.15), we obtain

$$\lim_n \frac{X_{[1,n]}(cN)}{n} = \lim_n \frac{X_{[-n,0]}(cN)}{n} = \gamma(c) + \frac{c}{\lambda}\ .$$

For $c \geqslant \tilde{c}$, we have $cN \geqslant \tilde{c}N$. We obtain by internal monotonicity and by monotonicity respectively :

1. $Z_{[-n,0]}(cN)$ is decreasing in $c \Longrightarrow \gamma(c)$ is decreasing in c.

2. $X_{[0,n]}(cN)$ is increasing in $c \Longrightarrow \gamma(c) + c/\lambda$ is increasing in c.

We deduce the existence of a constant $\gamma(0)$ defined by :

$$\lim_{c\to 0} \searrow \gamma(c) + \frac{c}{\lambda} = \gamma(0) = \lim_{c\to 0} \nearrow \gamma(c)\,. \tag{10.17}$$

The intuitive interpretation is that $\gamma(0)^{-1}$ is the throughput of the network when we saturate the input, i.e. when $A_n = 0, \forall n$. It is the maximal possible throughput.

Theorem 10.7.3. *Let $N = \{A_n, M_n, n \in \mathbb{Z}\}$ be a stationary ergodic point process. We set $\rho = \lambda\gamma(0)$. If $\rho > 1$, then $P(Z = +\infty) = 1$. If $\rho < 1$, then $P(Z < +\infty) = 1$ and $\{Z_{[0,n]}, n \in \mathbb{N}\}$ couples in finite time with the stationary sequence $\{Z \circ \theta^n\}$.*

Proof. The first part of the theorem is immediate. In fact relation (10.17) implies $\gamma(1) + 1/\lambda \geqslant \gamma(0)$. We have :

$$\left(\lim_n \frac{Z_{[-n,0]}}{n} = \gamma(1)\right) \geqslant \left(\gamma(0) - \frac{1}{\lambda} = \frac{\rho - 1}{\lambda}\right)\,.$$

Therefore $\rho - 1 > 0$ implies $P(Z = +\infty) = 1$. For a complete proof of the result, the reader is referred to [10]. □

Remark 10.7.4. For $\rho < 1$, Z is the smallest stationary regime for the response time of the system (which is defined as the time to the last activity under the restriction $[-\infty, 0]$ of N). Intuitively it is the stationary regime corresponding to an "empty" initial condition as it is the limit of the systems starting "empty" and fed up with the restrictions $[-n, 0]$ of N. In many cases, there will be multiple stationary regimes depending on the initial condition. A simple example of a monotone and separable open network having multiple stationary regimes is proposed in [6], p.83. It is a $G/G/2/\infty$ queue with a "shortest workload" allocation rule (see also Theorem 10.8.6).

10.7.2 General discrete event networks

For discrete event networks which are not open, there are no such second order results. The reason is the absence of internal monotonicity of the variables $Z_{[-n,0]} = X_{[-n,0]} - X^-_{[-n,0]}$. We illustrate the phenomenon on Figure 10.2 where we compare the case of a general network and the case of an open network.
For open and general networks, we consider successively the restrictions $[-n, 0]$ and $[-n-1, 0]$. In the open case, the internal monotonicity has been illustrated in Figure 10.2. In the general case, the variables X^- are internal variables, hence their value are modified when we go from the restriction $[-n, 0]$ to $[-n-1, 0]$. As a consequence, there is no internal monotonicity. On Figure 10.2, for the ease of comparison, we have assumed that $X^-_{[-n-1,0]}(-n) = X^-_{[-n-1,0]}(-n)$ (these quantities are defined up to an additive constant).

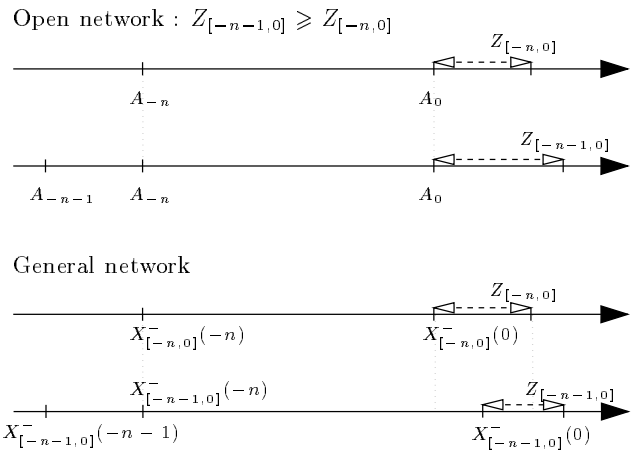


Figure 10.2: Loynes scheme for monotone-separable networks.

10.8 Application to Stochastic Operators

We propose in Sections 10.8.1 and 10.8.2 two very different approaches. They correspond to two different types of operators, see Remark 10.8.4. The first approach is directly based on the Loynes scheme. The second one uses fixed points results.

10.8.1 Monotonicity

Definition 10.8.1. *We say that the operator $T : \mathbb{R}^k \to \mathbb{R}^k$ has a minimal value if there exists $x_0 \in \mathbb{R}^k$ such that $\forall y \geqslant x_0$, $T(y) \geqslant x_0$.*

Let us consider a sequence of monotone operators $\{T_n, n \in \mathbb{Z}\}$. If all the operators have a common minimal value x_0, then we are able to construct a Loynes scheme, in the same way as in §10.6. In fact, we have $T_0(x_0) \geqslant x_0$ and $T_0 \circ T_{-1}(x_0) \geqslant T_0(x_0) \geqslant x_0$ using monotonicity. We obtain that

$$\exists Z \in (\mathbb{R} \cup \{+\infty\})^k, \lim_n T_0 \circ T_{-1} \circ \cdots \circ T_{-n}(x_0) = Z \,. \tag{10.18}$$

The main question is whether the limit Z is finite or not, the finite case being the interesting one. In particular, if we consider a sequence of monotone-homogeneous operators, then the limits $\overline{\gamma}$ and $\underline{\gamma}$ as defined in Proposition 10.3.4 exist. Because of the existence of the minimal value x_0, we have $\overline{\gamma} \geqslant \underline{\gamma} \geqslant 0$. If $\overline{\gamma} > 0$ then there exists i such that $Z_i = +\infty$ (the proof is immediate).

For this reason, it is usually not interesting to construct a Loynes scheme directly on the sequence of operators T_n. For example, in the case of the operator of the G/G/1 queue, see Equation (10.11), the Loynes scheme was not built on $(A_n, X_{[l,n]})'$ but on the differences $Z_{[l,n]} = X_{[l,n]} - A_n$. In order to generalize the construction, the good approach is to consider the operators T_n in a projective space.

We have already defined the projective space $\mathbb{PR}^k$ in Definition 10.3.7. The space $\mathbb{PR}^k$ is isomorphic to $\mathbb{R}^{k-1}$. Here are different possible ways to map $\mathbb{PR}^k$ onto $\mathbb{R}^{k-1}$. Let $i \in \{1, \ldots, k\}$, we define :

$$\pi_i : \ \mathbb{R}^k \longrightarrow \mathbb{R}^{k-1}, \quad \pi_i(x) = (x_1 - x_i, \ldots, x_{i-1} - x_i, x_{i+1} - x_i, \ldots, x_k - x_i)'$$
$$\phi_i : \ \mathbb{PR}^k \to \mathbb{R}^{k-1}, \quad \phi_i = \pi_i \circ \pi^{-1} ,$$

where π was defined in Definition 10.3.7. It is easy to verify that ϕ_i is defined without ambiguity and is bijective.

Definition 10.8.2. *Let $x \in \mathbb{R}^k$. We define $|x|_{\mathcal{P}} = \max_i x_i - \min_i x_i$. Let $u \in \mathbb{PR}^k$ (resp. $u \in \mathbb{R}^{k-1}$) and x be a representative of u, i.e. $\pi(x) = u$ (resp. $\pi_i(x) = u$) We define $|u|_{\mathcal{P}} = \max_i x_i - \min_i x_i$.*

The function $|.|_{\mathcal{P}}$ is a semi-norm on $\mathbb{R}^k$ as $|x|_{\mathcal{P}} = 0 \Rightarrow x_i = \lambda, \forall i$. On the other hand, it defines a norm on $\mathbb{PR}^k$ or $\mathbb{R}^{k-1}$. We call it the projective norm. We use the same notation for the semi-norm on $\mathbb{R}^k$ and the norms on $\mathbb{PR}^k$ and $\mathbb{R}^{k-1}$ in order not to carry too many notations. Form now on, we are going to work on $\mathbb{R}^{k-1}$ equipped with the projective norm. Without loss of generality, we will restrict our attention to π_1, ϕ_1. Working on $\mathbb{R}^{k-1}$ rather than on $\mathbb{PR}^k$ enables us to have a natural partial order. The projective norm is indeed compatible with the coordinatewise partial ordering on $\mathbb{R}^{k-1}$, i.e. $u, v \in \mathbb{R}^{k-1}, u \geqslant v \Rightarrow |u|_{\mathcal{P}} \geqslant |v|_{\mathcal{P}}$.

Let $T : \mathbb{R}^k \to \mathbb{R}^k$ be an homogeneous operator. We define

$$\tilde{T} : \ \mathbb{R}^{k-1} \to \mathbb{R}^{k-1} \quad \tilde{T}(u) = \pi_1(T(x)), \ x \in \pi_1^{-1}(u) .$$

Because of homogeneity, $\tilde{T}(u)$ is unambiguously defined. We can write, with abbreviated notations, $\tilde{T} = \pi_1 \circ T \circ \pi_1^{-1}$.

Lemma 10.8.3. *We consider an homogeneous operator $T : \mathbb{R}^k \to \mathbb{R}^k$ and the associated operator $\tilde{T} : \mathbb{R}^{k-1} \to \mathbb{R}^{k-1}$, satisfying the following assumptions:*

A. T is monotone.

B. $T(x)_1 - x_1$ is independent of $x \in \mathbb{R}^k$.

C. $\exists x_0$ such that $T(x_0)_1 - (x_0)_1 = \min_i (T(x_0)_i - (x_0)_i)$.

Under Assumption A, $\tilde{T}$ is non-expansive. Under Assumptions A + B, $\tilde{T}$ is monotone. Under Assumptions A+B+C, $\tilde{T}$ has minimal value $\tilde{x}_0 = \pi_1(x_0)$.

Proof. We consider $u, v \in \mathbb{R}^{k-1}$ verifying $u \geqslant v$. Let $x, y \in \mathbb{R}^k$ be such that $\pi_1(x) = u, \pi_1(y) = v$ and $x_1 = y_1$.

1. $A \Rightarrow \tilde{T}$ is non-expansive. The representatives x and y are such that $|u - v|_{\mathcal{P}} = |x - y|_{\mathcal{P}} = ||x - y||_\infty$. By monotonicity of T, we have $T(x) \geqslant T(y)$, hence $|T(x) - T(y)|_{\mathcal{P}} \leqslant ||T(x) - T(y)||_\infty$. By non-expansiveness of T (Theorem 10.3.2), we have $||T(x) - T(y)||_\infty \leqslant ||x - y||_\infty$. We conclude that :

$$\begin{aligned} |\tilde{T}(u) - \tilde{T}(v)|_{\mathcal{P}} = |T(x) - T(y)|_{\mathcal{P}} &\leqslant ||T(x) - T(y)||_\infty \\ &\leqslant ||x - y||_\infty = |u - v|_{\mathcal{P}} . \end{aligned}$$

2. $A+B \Rightarrow \tilde{T}$ is monotone. Let the representatives x and y verify $x_1 = y_1$. Hence by Assumption B, we have $T(x)_1 = T(y)_1$. We conclude that $T(x) \geqslant T(y) \Rightarrow \tilde{T}(u) \geqslant \tilde{T}(v)$.

3. $A + B + C \Rightarrow \tilde{T}$ has minimal value $\tilde{x}_0 = \pi_1(x_0)$. We have

$$\begin{aligned} \tilde{T}_0(\tilde{x}_0)_i &= T(x_0)_i - T(x_0)_1 = T(x_0)_i - (x_0)_i + (x_0)_i - T(x_0)_1 \\ &\geqslant T(x_0)_1 - (x_0)_1 + (x_0)_i - T(x_0)_1 = (\tilde{x}_0)_i \,. \end{aligned}$$

We conclude with the monotonicity of $\tilde{T}$ that $\forall y \in \mathbb{R}^{k-1}, y \geqslant \tilde{x}_0 \Rightarrow \tilde{T}(y) \geqslant \tilde{x}_0$.

□

The operator $\tilde{T}$ is not homogeneous in general. Hence the conditions ensuring monotonicity and non-expansiveness are not the same (to be compared with Theorem 10.3.2).

Remark 10.8.4. Assumption B. can be easily weakened and replaced by :

$$B'.\ \forall x, y \in \mathbb{R}^k,\ x_1 - y_1 = \min_i x_i - y_i \Rightarrow T(x)_1 - T(y)_1 = \min_i T(x)_i - T(y)_i \,.$$

In Lemma 10.8.3, we have presented the assumptions which appear naturally in physical systems. In particular, Assumption B is verified when the first coordinate of T is the dater of an exogenous arrival process. Assumption C is verified if the other coordinates of T correspond to events which are induced by the arrivals (hence occur later on). It was the case for the operator associated with the $G/G/1/\infty$ queue, see Equation (10.11). In that example, the minimal value was $e = (0, \ldots, 0)'$.
These assumptions are of course restrictive. Roughly speaking, they will apply only to some operators associated with 'open systems'. For operators associated with 'closed systems', the conditions and results of Section 10.8.2 are more appropriate.

Theorem 10.8.5. *Let $\{T_n, n \in \mathbb{N}\}$ be a stationary and ergodic sequence of homogeneous random operators on $\mathbb{R}^k$ and $\{\tilde{T}_n, n \in \mathbb{N}\}$ the associated sequence on $\mathbb{R}^{k-1}$. We assume that Assumptions A,B and C of Lemma 10.8.3 are verified with probability 1 by the operators $\{T_n\}$ (in particular they have a constant minimal value x_0). We set $\tilde{x}_0 = \pi_1(x_0)$. Then the limit $Z = \lim_n \tilde{T}_0 \circ \cdots \circ \tilde{T}_{-n}(\tilde{x}_0)$ exists and verifies $P\{Z < +\infty\} = 0$ or 1. Furthermore Z is a stationary solution, i.e. $Z(\theta\omega) = \tilde{T}_1(Z(\omega))$. When $P\{Z < +\infty\} = 1$, the sequence $\{T_n \circ \cdots \circ T_1(x_0)\}$ couples in finite time with the stationary sequence $\{Z \circ \theta^n\}$.*

Proof. It is exactly similar to the one of Loynes Theorem 10.6.1. □

The main difficulty is often to prove the finiteness of Z. Moreover, when finite, Z is usually not the unique stationary solution. Indeed, we have that $\forall \lambda \in \mathbb{R}$, $\tilde{x}_0 + \lambda \vec{1}$ is a minimal value for the operators $\tilde{T}_n$. Hence by Theorem 10.8.5, the limits

$$Z^\lambda = \lim_n \tilde{T}_0 \circ \cdots \circ \tilde{T}_{-n}(\tilde{x}_0 + \lambda \vec{1})$$

exist and are stationary solutions. The variables Z^λ are increasing in λ by monotonicity of $\tilde{T}_n$. Hence we can define the limit

$$Z^\infty = \lim_{\lambda \to +\infty} Z^\lambda \,. \tag{10.19}$$

Next theorem was originally proved by Brandt for a special operator associated with the queue $G/G/k/\infty$.

Theorem 10.8.6 (Brandt [31]). *We have $P\{Z^\infty < +\infty\} = 0$ or 1. If $P\{Z^\infty < +\infty\} = 1$, then Z^∞ is the maximal finite stationary solution, i.e $Z(\theta\omega) = \tilde{T}_1(Z(\omega))$ and*

$$Y(\theta\omega) = \tilde{T}_1(Y(\omega)),\ P\{Y < +\infty\} = 1 \Rightarrow P\{Z^\infty \geqslant Y\} = 1\,.$$

Proof. The essential ingredient is the non-expansiveness of $\tilde{T}_n$. For more details, the reader is referred to [31] or [32], Theorem 1.3.2. □

Remark 10.8.7. The results presented in this section §10.8.1 are just a specialization to operators of finite dimension of more general results. Let $(E, \mathcal{E})$ be a Polish space (complete separable metric space) equipped with its Borel σ-field. We consider $\{\phi_n, n \in \mathbb{Z}\}$ a stationary and ergodic sequence of measurable random functions $\phi_n : E \times \Omega \to E$. The recursive equations $x(n+1) = \phi_n(x(n)), x(0) = x_0$ define a Stochastic Recursive Sequence, following the terminology of Borovkov [22]. If the functions ϕ_n are monotone and verify $\phi_n(x_0) \geqslant x_0$ then the results of Theorem 10.8.5 hold (replace just T_n by ϕ_n). If we assume furthermore that the functions ϕ_n are non-expansive (with respect to the metric of E) then the results of Theorem 10.8.6 hold. For a detailed presentation of this framework, see [32] [25].

10.8.2 Fixed point

In this section, we will see a rather different use of Loynes backward construction.
Here is a result generalizing Proposition 10.3.9. The proof of $A \Rightarrow B$ in Prop. 10.3.9 was using only the continuity of the operator T. In fact, using the non-expansiveness of T, we can get stronger results.

Theorem 10.8.8 (Weller [138], Sine [130]). *Let C be a compact of $\mathbb{R}^k$. We consider an operator $T : C \to C$, non-expansive with respect to the sup-norm $||.||_\infty$. Then we have :*

$$\forall x \in C, \exists p \in \mathbb{N}, \exists u \in C\ :\ \lim_{n\to\infty} T^{np}(x) = u \text{ and } T^p(u) = u\,. \tag{10.20}$$

Corollary 10.8.9. *Let T be defined as in Theorem 10.8.8. We assume that $\forall n \geqslant 1$, T^n has a unique fixed point u. Then $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ such that*

$$\forall n \geqslant N\ :\ \sup_{x\in C} ||T^n(x) - u||_\infty \leqslant \varepsilon\,. \tag{10.21}$$

In other words, there is uniform convergence of T^n to u.

Proof. Let us prove first that T^n converges simply to u. Let x belong to C. As u is the unique fixed point of the powers of T, we obtain by application of Theorem 10.8.8 :

$$\forall x \in C,\ \exists p \in \mathbb{N},\ \forall \varepsilon > 0,\ \exists M(x,\varepsilon) \in \mathbb{N}, \forall n \geqslant M(x,\varepsilon)\ :\ ||T^{np}(x) - u||_\infty \leqslant \varepsilon\,.$$

By non-expansiveness, we have $\left(||T \circ T^{np}(x) - T(u)||_\infty = ||T^{np+1}(x) - u||_\infty\right) \leqslant ||T^{np}(x) - u||_\infty$ and by induction, $\forall q \in \mathbb{N}, ||T^{np+q}(x) - u||_\infty \leqslant ||T^{np}(x) - u||_\infty$. It implies that

$$\forall x \in C,\ \forall \varepsilon > 0,\ \exists N(x,\varepsilon) \in \mathbb{N}, \forall n > N(x,\varepsilon),\quad ||T^n(x) - u||_\infty \leqslant \varepsilon\,.$$

We are now ready to prove that the convergence is uniform. Let us denote by $\mathcal{B}(x,\varepsilon)$ the open ball of center x and radius ε for the sup-norm. Using non-expansiveness, we have that $\forall y \in \mathcal{B}(x,\varepsilon), \forall n \geqslant N(x,\varepsilon)$,

$$||T^n(y) - u||_\infty \leqslant ||T^n(y) - T^n(x)||_\infty + ||T^n(x) - u||_\infty \leqslant 2\varepsilon \,. \tag{10.22}$$

Using Borel-Lebesgue's characterization of compact sets, there exists a finite number of points x_i such that $C \subset \bigcup_i \mathcal{B}(x_i,\varepsilon)$. Using Equation (10.22), we obtain :

$$\forall \varepsilon > 0,\ \forall n \geqslant \max_i N(x_i,\varepsilon),\ \forall x \in C\ :\quad ||T^n(x) - u||_\infty \leqslant 2\varepsilon \,.$$

This completes the proof. □

We are now ready to prove the main theorem of this section which generalizes the results of Chapter 6.

Theorem 10.8.10. *Let $\{T_n, n \in \mathbb{N}\}$ be a stationary ergodic sequence of monotone homogeneous random operators on $\mathbb{R}^k$ and $\{\tilde{T}_n\}$ the associated sequence on $\mathbb{R}^{k-1}$. We assume that there exists a deterministic monotone homogeneous operator S on $\mathbb{R}^k$ ($\tilde{S}$ on $\mathbb{R}^{k-1}$) such that*

i. $\tilde{S}$ is bounded, i.e. there exists a compact K of $\mathbb{R}^{k-1}$ such that $\mathrm{Im}(\tilde{S}) \subset K$.

ii. $\forall n \geqslant 1$, $\tilde{S}^n$ has a unique fixed point, u.

iii. There exists a deterministic constant l such that $\tilde{S}$ belongs to the support of the random operator $\tilde{T}_l \circ \cdots \circ \tilde{T}_1$ and $\forall n > 0, \tilde{S}^n$ belongs to the support of $\tilde{T}_{nl} \circ \cdots \circ \tilde{T}_1$, with the following precise meaning :

$$\forall \varepsilon > 0,\quad P\{\sup_{x \in \mathbb{R}^{k-1}} |\tilde{T}_l \ldots \tilde{T}_1(x) - \tilde{S}(x)|_{\mathcal{P}} \leqslant \varepsilon\} > 0,$$
$$P\{\sup_{x \in \mathbb{R}^{k-1}} |\tilde{T}_{nl} \ldots \tilde{T}_1(x) - \tilde{S}^n(x)|_{\mathcal{P}} \leqslant \varepsilon\} > 0 \,.$$

Then $\forall x \in \mathbb{R}^{k-1}$, $\tilde{x}(n) = \tilde{T}_{n-1} \circ \cdots \circ \tilde{T}_0(x)$ converges weakly to a unique stationary distribution.

Proof. We first prove the theorem when replacing Assumptions *iii.* by the stronger assumption :

iv. $\exists l$ s.t. $P\{\tilde{T}_l \circ \cdots \circ \tilde{T}_1 = \tilde{S}\} > 0$ and $\forall n > 0$, $P\{\tilde{T}_{nl} \circ \cdots \circ \tilde{T}_1 = \tilde{S}^n\} > 0$.

For $x \in \mathbb{R}^{k-1}$, we define the variables :

$$Z_{-n,0}(x) = \tilde{T}_0 \circ \cdots \circ \tilde{T}_{-n}(x) = \tilde{x}(n,x) \circ \theta^{-n} \,. \tag{10.23}$$

We now prove that $Z_{-n,0}(x)$ admits *P.a.s.* a limit which is independent of x.

The compact K of Assumption $i.$ is stable by $\tilde{S}$ and by Assumption $ii.$, there is a unique fixed point $u \in \mathbb{R}^{k-1}$ for the powers of $\tilde{S}$. From Lemma 10.8.3, $\tilde{S}$ is non-expansive with respect to the projective norm. Hence Corollary 10.8.9 can be applied to $\tilde{S}$ on $(\mathbb{R}^{k-1}, |.|_{\mathcal{P}})$. It implies

$$\forall \varepsilon > 0,\ \exists N(\varepsilon),\ \forall n \geqslant N(\varepsilon),\ \forall x \in \mathbb{R}^{k-1},\quad |\tilde{S}^n(x) - u|_{\mathcal{P}} \leqslant \varepsilon\ . \tag{10.24}$$

We define the random variables

$$\forall \varepsilon > 0,\ M(\varepsilon) = \min\{n \geqslant N(\varepsilon) l \mid \tilde{T}_{-n} \circ \cdots \circ \tilde{T}_{-n-N(\varepsilon)l+1} = \tilde{S}^{N(\varepsilon)}\}\ , \tag{10.25}$$

where $N(\varepsilon)$ and l are defined in Equation (10.24) and in Assumption $iv.$ respectively. Assumption $iv.$ also implies that $P\{M(\varepsilon) < +\infty\} > 0$. We obtain

$$P\{M(\varepsilon) < +\infty\} = 1\ , \tag{10.26}$$

in exactly the same way as we obtained $P\{N_1 < +\infty\} = 1$ in the proof of Theorem 10.3.10.
Let us fix $\varepsilon = 1$. We define the events $\mathcal{A}_n = \{M(1, \omega) = n\}$ which form a countable partition of Ω.
Let us work for a moment on the event $\mathcal{A} = \mathcal{A}_m$ for a given integer m. Let us consider the variables $Z_{-n,-m}(x) = \tilde{T}_{-m} \circ \ldots \tilde{T}_{-n}(x),\ n > m$. We have

$$\forall n \geqslant m + N(1)l,\ Z_{-n,-m}(x) = \tilde{S}^{N(1)}(\tilde{T}_{-m-N(1)^l} \circ \cdots \circ \tilde{T}_{-n}(x))\ . \tag{10.27}$$

Hence, on $\mathcal{A}_m$, the image of $Z_{-n,-m}$ is included in the closed ball of center u and radius 1 (Equation (10.24)) that we denote by $K(1)$,

$$\forall n \geqslant m + N(1)l,\ \mathrm{Im}(Z_{-n,-m}) \subset K(1)\ . \tag{10.28}$$

We consider the sequence of random variables $\{M(1/i), i \in \mathbb{N}\}$. By definition of the variables $M(\varepsilon)$, (10.25), the sequence $M(1/i)$ is increasing in i in particular $M(1/i) \geqslant M(1)$. We have, for all $n \geqslant M(1/i) + N(1/i)l$ (note that we consider the variables Z with respect to an unchanged ending point $-m$).

$$\begin{aligned} &Z_{-n,-m}(x) = \tilde{T}_{-m} \circ \cdots \circ \tilde{T}_{-n}(x) \\ &= \tilde{T}_{-m} \circ \cdots \circ \tilde{T}_{-M(1/i)+1} \circ\ \tilde{S}^{N(1/i)}\ \circ \tilde{T}_{-M(1/i)-N(1/i)l} \circ \cdots \circ \tilde{T}_{-n}(x)\ . \end{aligned}$$

Using Equation (10.24), we have that $\tilde{S}^{N(1/i)} \circ \tilde{T}_{-M(1/i)-N(1/i)l} \circ \cdots \circ \tilde{T}_{-n}(x)$ is included in the closed ball of center u and radius $1/i$. Using the non-expansiveness of the operators, we obtain the existence of a compact set, denoted $K(1/i)$ such that

$$\forall n \geqslant M(\frac{1}{i}) + N(\frac{1}{i})l,\ \ \mathrm{Im}(Z_{-n,-m}) \subset K(\frac{1}{i})\ . \tag{10.29}$$

We have built a decreasing sequence of compact sets $K(1/i)$ whose radius goes to zero. By a classical theorem on decreasing sequences of compact sets (Borel-Lebesgue Theorem), the intersection of the sets $K(1/i)$ is a single point. It means precisely that the limit of $Z_{-n,-m}(x),\ n \to +\infty$, exists and is independent of x. We define the following notations

$$\forall \omega \in \mathcal{A}_m,\ \forall x \in \mathbb{R}^{k-1},\ \lim_{n\to\infty} Z_{-n,-m}(x) = Z_{\infty,m},\ Z = \tilde{T}_0 \circ \cdots \circ \tilde{T}_{-m+1}(Z_{\infty,m})\ .$$

It is straightforward to prove that $Z = \lim_{n\to+\infty} Z_{-n,0}(x)$. By applying the same construction to all the events $\mathcal{A}_m, m \in \mathbb{N}$, we prove the a.s. existence of $Z = \lim_n Z_{-n,0}(x)$, the limit being independent of x. By analogy with §10.6, we call Z the Loynes variable.

We are now going to prove the existence of the Loynes variable Z under the weaker Assumption *iii*.
We define the random variables $N(\varepsilon)$ as previously, (10.24). On the other hand, the definition of the variables $M(\varepsilon)$ is modified

$$\forall \varepsilon,\ M(\varepsilon) = \min\{n \geqslant N(\varepsilon)l \mid \sup_{x\in\mathbb{R}^{k-1}} |\tilde{T}_{-n+N(\varepsilon)l} \ldots \tilde{T}_{-n}(x) - \tilde{S}^{N(\varepsilon)}(x)|_{\mathcal{P}} \leqslant \varepsilon\}\ . \tag{10.30}$$

From Assumption *iii*. and the Ergodic Lemma 10.2.1, we obtain $P\{M(\varepsilon) < +\infty\} = 1$.
We define the variable $M(1)$, then the partition $\mathcal{A}_n$, the event $\mathcal{A}$ and the variables $M(1/i)$ as before. We define the variables :

$$\hat{Z}^i_{-n,-m}(x) = \tilde{T}_{-m} \ldots \tilde{T}_{-M(1/i)+1} \circ \left(\tilde{S}^{N(1/i)}\right) \circ \tilde{T}_{-M(1/i)-1-N(1/i)l} \ldots \tilde{T}_{-n}(x) \tag{10.31}$$

There exists a sequence of compacts $\hat{K}(1/i)$ of radius $1/i$ such that (see the first part of the proof)

$$\forall n \geqslant M(1/i) + N(1/i)l,\ \mathrm{Im}(\hat{Z}^i_{-n,-m}) \subset \hat{K}(1/i)\ . \tag{10.32}$$

From the definition of $M(1/i)$, Equation (10.30), we get

$$\forall n \geqslant M(1/i) + N(1/i)l, \forall x \in \mathbb{R}^{k-1},\ |Z_{-n,-M(1/i)}(x) - \hat{Z}^i_{-n,-M(1/i)}(x)|_{\mathcal{P}} \leqslant \frac{1}{i}\ .$$

Using non-expansiveness, we obtain

$$\forall n \geqslant M(1/i) + N(1/i)l, \forall x \in \mathbb{R}^{k-1},\ |Z_{-n,-m}(x) - \hat{Z}^i_{-n,-m}(x)|_{\mathcal{P}} \leqslant \frac{1}{i}\ .$$

We conclude that $\forall n \geqslant M(1/i) + N(1/i)l, \forall x, y \in \mathbb{R}^{k-1}$

$$\begin{aligned} |Z_{-n,-m}(x) - \ \ Z_{-n,-m}(y)|_{\mathcal{P}} &\leqslant |Z_{-n,-m}(x) - \hat{Z}^i_{-n,-m}(x)|_{\mathcal{P}} + \\ &|\hat{Z}^i_{-n,-m}(x) - \hat{Z}^i_{-n,-m}(y)|_{\mathcal{P}} + |\hat{Z}^i_{-n,-m}(y) - Z_{-n,-m}(y)|_{\mathcal{P}} \leqslant \frac{3}{i}\ . \end{aligned}$$

Hence there exists a sequence of compacts $K(1/i)$ of radius $3/i$ such that $\forall n \geqslant M(1/i) + N(1/i)l, \mathrm{Im}(Z_{-n,-m}) \subset K(1/i)$. We conclude as in the first part of the proof.

Our aim is now to prove that we have weak convergence of the process $\tilde{x}(n) = \tilde{T}_n \circ \cdots \circ \tilde{T}_0(x(0))$ to the stationary distribution of Z. We consider a function $f : \mathbb{R}^k \to \mathbb{R}$, continuous and bounded. We have, using the stationarity of $\{\tilde{T}_n\}$

$$\begin{aligned} E\,(\, f(x(n, x(0)))\,) &= E(\, f(\tilde{T}_{n-1} \circ \cdots \tilde{T}_0(x(0)))\,) \\ &= E(\, f(\tilde{T}_0 \circ \cdots \tilde{T}_{-n+1}(x(0)))\,) \xrightarrow{n} E\, f(Z) \end{aligned} \tag{10.33}$$

The convergence in (10.33) is obtained from Lebesgue's dominated convergence theorem (f is bounded). It proves weak convergence. □

Remark 10.8.11. It would be nice to replace Assumption *iii.* by the following weaker Assumption

$$\begin{aligned} v. \qquad \forall \varepsilon > 0,\ \forall K \text{ compact}\ , \quad & P\{\sup_{x \in K} |\tilde{T}_l \ldots \tilde{T}_1(x) - \tilde{S}(x)|_{\mathcal{P}} \leqslant \varepsilon\} > 0, \\ & P\{\sup_{x \in K} |\tilde{T}_{nl} \ldots \tilde{T}_1(x) - \tilde{S}^n(x)|_{\mathcal{P}} \leqslant \varepsilon\} > 0\ . \end{aligned}$$

Assumption $v.$ means precisely that $\tilde{S}$ is in the support of $\tilde{T}_0$ for the topology of weak convergence on the functional space $C_0(\mathbb{R}^{k-1}, \mathbb{R}^{k-1})$ (continuous functions of $\mathbb{R}^{k-1}$).
However, Theorem 10.8.10 is not true under Assumption v. Here is a counter-example. We consider $a, b \in \mathbb{R}^+$ and we define the monotone homogeneous operators on $\mathbb{R}^2$:

$$T_A(x) = \begin{pmatrix} x_1 \\ x_2 + a \end{pmatrix} \tag{10.34}$$

$$\forall i \in \mathbb{N}^+,\ T_{B_i}(x) = \begin{pmatrix} x_1 \\ \max(x_2 - ib, x_1) \end{pmatrix} \tag{10.35}$$

We consider a sequence of i.i.d. random operators $\{T_n, n \in \mathbb{N}\}$ with the following distribution :

$$P\{T_0 = T_A\} = \frac{1}{2},\ P\{T_0 = T_{B_i}\} = \frac{1}{2^{i+1}}, i \in \mathbb{N}^+\ .$$

We define the monotone homogeneous operator $S : \mathbb{R}^2 \to \mathbb{R}^2,\ S(x) = (x_1, x_1)'$. It is clear that $\tilde{S}$ verifies the Assumptions *i.* and *ii.* as $\tilde{S}$ is constant. Let K be a compact set of $\mathbb{R}$ and n be such that $K \subset [-n, n]$. We obtain immediately that $\forall x \in K,\ \tilde{T}_{B_i}(x) = \tilde{S}(x)$ as soon as $ib \geqslant n$. Hence $\tilde{S}$ verifies also Assumption v.
The description of the process $\tilde{x}(n) = \tilde{T}_{n-1} \circ \cdots \circ \tilde{T}_0(0)$ is very easy. It is a random walk on the real line with an absorbing barrier at 0. The drift of the random walk is

$$\delta = \frac{a}{2} - \sum_{i=1}^{\infty} \frac{ib}{2^{i+1}} = \frac{a}{2} - b\ .$$

We conclude that the process $\tilde{x}(n)$ is transient if $a > 2b$ which provides the announced counter-example.

Practically speaking, the main difficulty consists in finding a deterministic operator S verifying the assumptions of Theorem 10.8.10. We discuss this point for some specific models in §10.9.1.

10.9 Models Entering the Framework

10.9.1 Stochastic operators

Let $\mathcal{A}$ and $\mathcal{B}$ be two arbitrary sets. We define applications ($\mathbb{M}_k$ denotes the set of matrices of dimension $k \times k$)

$$P : \mathcal{A} \times \mathcal{B} \to \mathbb{M}_k(\mathbb{R}),\ A : \mathcal{A} \times \mathcal{B} \to \mathbb{M}_k(\mathbb{R} \cup \{-\infty, +\infty\}) ,$$

where the matrices $P(\alpha, \beta)$ are "markovian", i.e. verify

$$\forall i \in \{1, \ldots, k\},\ p_{ij}(\alpha, \beta) \geqslant 0,\ \sum_{j=1}^{k} p_{ij}(\alpha, \beta) = 1 . \tag{10.36}$$

Let us consider the following "(min,max,+,×)" operator

$$x \in \mathbb{R}^k, i \in \{1, \ldots, k\},\ T(x)_i = \inf_{\alpha \in \mathcal{A}} \sup_{\beta \in \mathcal{B}} \sum_{j=1}^{k} p_{ij}(\alpha, \beta)\ (x_j + a_{ij}(\alpha, \beta)) . \tag{10.37}$$

Equation (10.37) arises in stochastic control of dynamic games, see for example [19]. If $T(x)_i$ is finite ($\forall x \forall i$) then it defines a monotone-homogeneous operator. For example, let us prove homogeneity. We have for $x \in T^k, \lambda \in \mathbb{R}$,

$$\begin{aligned} T(x + \lambda\vec{1})_i &= \inf_{\alpha} \sup_{\beta} \sum_{j=1}^{k} p_{ij}(\alpha, \beta)\ (x_j + \lambda + a_{ij}(\alpha, \beta)) \\ &= \inf_{\alpha} \sup_{\beta} (\sum_{j=1}^{k} p_{ij}(\alpha, \beta)\lambda) + \sum_{j=1}^{k} p_{ij}(\alpha, \beta)\ (x_j + a_{ij}(\alpha, \beta)) = \lambda + T(x)_i . \end{aligned}$$

The following representation theorem provides a precise idea of the degree of generality of the class of monotone-homogeneous operators.

Theorem 10.9.1 (Kolokoltsov [98]). *Let $T : \mathbb{R}^k \to \mathbb{R}^k$ be a monotone-homogeneous operator. Then it can be represented in the form of Equation (10.37).*

The next lemma which is based on this representation, is proved in [98]. It can be coupled with Theorem 10.8.10 to obtain second order results for some stochastic operators.

Lemma 10.9.2. *Let $T : \mathbb{R}^k \to \mathbb{R}^k$ be a monotone-homogeneous operator, written in the form of Equation (10.37). Let us assume that*

$$\exists \eta > 0 :\ \forall i, j \exists l : \forall \alpha, \beta,\ p_{il}(\alpha, \beta) > \eta,\ p_{jl}(\alpha, \beta) > \eta .$$

Then the operators $T^n, n \in \mathbb{N}$, have a unique generalized fixed point.

From the point of view of applications, the interesting case is when the sets $\mathcal{A}$ and $\mathcal{B}$ are finite. Here are some specializations of Equation (10.37).

(+,×) linear systems The operator T is just a markovian matrix P, see Equation (10.36). We have $T(x) = Px$ (matrix-vector multiplication in the usual algebra). Matrix P can be interpreted as the matrix of transition probabilities of a Markov Chain (MC) having state space $\{1, \ldots, k\}$. The most interesting operator for a MC is $S(y) = yP$ where y is a row vector. It is well known that the limit of $S^n(y), y \geqslant 0, \sum_i y_i = 1$ is the stationary distribution of the MC. But the operator $T(x) = Px$ is also interesting from the point of view of applications. It appeared in [54] to model the problem of reaching agreement on subjective opinions. More generally, it has been studied as a special case of the general theory of products of non-negative matrices, see for example [127], Chapter 4.6.

For any markovian matrix P, we have $T(\vec{1}) = P\vec{1} = \vec{1}$. Hence the vector $\vec{1}$ is a generalized fixed point (Def. 10.3.8) of operator T. By application of the Perron-Frobenius Theorem, it is the only one. Hence, applying the ergodic results of this paper to a stochastic sequence of matrices P_n, is going to yield trivial results (the convergence of $\pi(P_n \ldots P_0 x)$ to $\pi(\vec{1})$). In fact much stronger results are known for such models. The necessary and sufficient conditions of convergence of $\pi(P_n \ldots P_0 x)$ to $\pi(\vec{1})$, are known for a general sequence of matrices P_n, without any stochastic assumptions, see [127], Th. 4.18.

(max,+) linear systems Such operators have the following form

$$x \in \mathbb{R}^k, i \in \{1, \ldots, k\}, \quad T(x)_i = \max_j (x_j + a_{ij}) \ , \tag{10.38}$$

$$T(x) = A \otimes x \ . \tag{10.39}$$

Equation (10.38) can be interpreted as a matrix-vector product in the (max,+) algebra. Equation (10.39) is simply a rewriting of Equation (10.38) using (max,+) notations. The (min,+) linear case boils down to the (max,+) case by switching to operator $-T$.
Such systems appear in many domains of applications, under various forms. For example (without any kind of exhaustiveness)

- Computer science : parallel algorithms, shared memory systems, PERT graphs, see [136] or [72].
- Queueing theory : $G/G/1/\infty$ queue (see §10.6), queues in series, queues in series with blocking, fork-join networks [8].
- Operations research and manufacturing : Job-shop models, event graphs (a subclass of Petri nets), see [44], [85] and [8].
- Economy or control theory : dynamic optimization, see [142].
- Physics of crystal structures : Frenkel-Kontorova model, see [80].

Among the very large and complete literature on the theoretical aspects of deterministic (max,+) systems, let us quote only [8] [107] and the references therein. As far as we know, the first references on stochastic (max,+) linear systems are [46] and [115]. Thanks to the rich deterministic theory, Theorems 10.3.5, 10.8.10 become very operational for (max,+) systems. The different

assumptions in these theorems can be interpreted as properties of the underlying graph structure of the model. For more details, see [105].

(min,max,+) linear systems These systems can be represented in one of the following dual forms. We use the symbol $\otimes$ for the (max,+) matrix-vector product, see (10.39), and the symbol $\odot$ for the (min,+) matrix-vector product.

$$x \in \mathbb{R}^k, \quad T(x) = \min\left(A_1 \otimes x, A_2 \otimes x, \ldots, A_l \otimes x\right) ,$$
$$T(x) = \max\left(B_1 \odot x, B_2 \odot x, \ldots, B_p \odot x\right) .$$

Here are some domains of application where such systems appear

- Minimax control in dynamic game theory, see [19].
- Study of timed digital circuits, see [82]. The (min,max) structure arises from the (and,or) operations of logical circuits.
- Queueing theory. G/G/s/∞ file, resequencing file, see for example [6]. Parallel processing systems [17] : there are k processors. A customer requires to use concurrently p out of the k processors to be executed.
- Motion of interfaces in particle systems [61]. As an illustration, let us describe a little bit more precisely a special case known as the marching soldier model. There is a row of k soldiers which advance in the same direction. In order to try to keep a common pace, they adopt the following strategy. At regular instants of time, each soldier checks the position of his right and left neighbours. He advances of 1 if they both are ahead of him and stays at the same position otherwise. Let $x \in \mathbb{R}^k$ denote the position of the soldiers at instant 0. Their position at instant 1, will be (with the convention $x_0 = x_{k+1} = +\infty$)

$$T(x)_i = \max\left(\min(x_{i-1}, x_i, x_{i+1}) + 1, x_i\right) .$$

The study of deterministic (min,max,+) systems (existence of generalized fixed points, projective boundedness,...) has been considered in several papers [114] [81]. However, it is far from being complete. For this reason, the only references on stochastic (min,max,+) systems concern first order results [61] [91].

(max,+,$\times$) linear systems These systems can be represented under the following form

$$x \in \mathbb{R}^k, \; T(x)_i = \max_{\alpha \in \mathcal{A}} \sum_{j=1}^{k} p_{ij}(\alpha)\left(x_j + a_i(\alpha)\right) . \tag{10.40}$$

Equation (10.40) appears in many domains of applications like operational research, management science and engineering. It is in fact one of the optimality equation of stochastic[3] dynamic

[3]The term stochastic refers here to the markovian interpretation of matrices $P(\alpha)$. According to our terminology, Equation (10.40) is that of a deterministic operator.

programming in discrete time, on a finite state space and with undiscounted rewards. A controller observes a system which evolves in a state space $\{1,\ldots,k\}$. The set of possible decisions for the controller is $\mathcal{A}$. Under decision $\alpha \in \mathcal{A}$, the system evolves from a state i to a state j according to the transition probabilities $p_{ij}(\alpha)$. Also, under decision $\alpha \in \mathcal{A}$, there is an immediate reward for being originally in state i which is $a_i(\alpha)$. It is well known that the optimal decision and the reward vector are obtained as $\lim_n T^n(x)$, see for example [140], Chapter 3.2. There is a very important literature on deterministic operators of type (10.40), see [124] or [140] and the references there. Next theorem is classical, for a proof see for example [140] Chapter 4.3.

Theorem 10.9.3. *Let T be an operator verifying Equation (10.40). A sufficient condition for the existence of a unique generalized fixed point for T is :*
$\forall \alpha \in \mathcal{A}$, matrix $P(\alpha)$ is ergodic, i.e. the graph of the non-zero terms of $P(\alpha)$ is strongly connected and aperiodic.

Remark 10.9.4. A (max,+) system can be viewed as a (max,+,$\times$) system with $\mathcal{A} = \{1,\ldots,k\}$ and $P(\alpha)$ is defined by $P_{ij}(\alpha) = 1$ if $j = \alpha$ and $P_{ij}(\alpha) = 0$ otherwise. Such matrices do not verify the assumption of Theorem 10.9.3.

The theorems presented in this paper, when coupled with results like Theorem 10.9.3 can be used in an efficient way for systems verifying (10.40) when the rewards $a(\alpha)$ and/or the transition matrices $P(\alpha)$ become random. The authors do not know of any reference on the subject.

10.9.2 Discrete event networks

We are now going to review some classes of discrete event networks. We restrict our attention to systems which can not be modeled as monotone-homogeneous operators. The references that are quoted are only the ones using the monotone separable framework or similar approaches.

- Precedence constraints models. Their study has been motivated by database systems. Different variations are considered in [13] [18] [51].
- Polling models. A wide class of polling models with general routing policies and stationary ergodic inputs enters the monotone separable framework, see [39].
- Free choice Petri nets. Event graphs, which are represented as (max,+) linear operators, see §10.9.1, or Jackson networks, see below, are subclasses of Free choice Petri nets. Free choice Petri nets enter the monotone separable framework, see [11] [9] [**?**].

Let us detail two of these models. First we propose a simple example of precedence constraint system and second Jackson networks.

Precedence constraints models There is a stream of customers $j(n), n \in \mathbb{N}$. Each customer $j(n)$ has a service time requirement $t(n)$ and precedence constraints under the form of a list $L(n)$ of customers. More precisely, we have $L(n) = \{j(i_1), j(i_2), \ldots j(i_{l_n})\}$ with $n > i_1 > i_2 > \cdots > i_{l_n} \geqslant 0$. Job j_n starts its execution as soon as all the customers of the list $L(n)$ have completed their execution. The execution of customer $j(n)$ takes $t(n)$ units of time.
Let us distinguish two cases.

1. We assume that the length of the precedence list is uniformly bounded by k, i.e. $\forall n \in \mathbb{N}, l_n \leqslant k$. We define the vector $x(n) \in \mathbb{R}^k$ such that $x(n)_i$ is the instant of completion of customer $j(n-i)$. From the dynamic described above, we have $x(n+1) = T_n(x(n))$, where the operator $T_n : \mathbb{R}^k \to \mathbb{R}^k$ is defined as follows

$$\begin{cases} T_n(x)_1 = \max_{\{i \mid j(n-i) \in L(n)\}} x_i + t(n) \\ T_n(x)_i = x_{i-1}, \ i = \{2, \ldots, k\} \end{cases}$$

   This operator is monotone homogeneous. It is in fact a (max,+) linear system, see §10.9.1.

2. Let us assume now that the length l_n is not uniformly bounded. It is not possible to describe the system as an operator of finite dimension. Let $X_{[1,n]}$ be the last instant of completion of one of the customers $j(i), i \in \{1, \ldots, n\}$. It is easy to verify that $X_{[1,n]}$ verifies the properties of the monotone-separable framework for discrete event networks, see §10.4.

In both cases, when $\{t(n), L(n), n \in \mathbb{N}\}$ forms a stationary ergodic sequence of random variables, we can apply the ergodic theorems presented in this paper.

Jackson networks Jackson networks were introduced by Jackson in [90]. It is a queueing network with I nodes, where each node is a single server FIFO queue, see Figure 10.3.

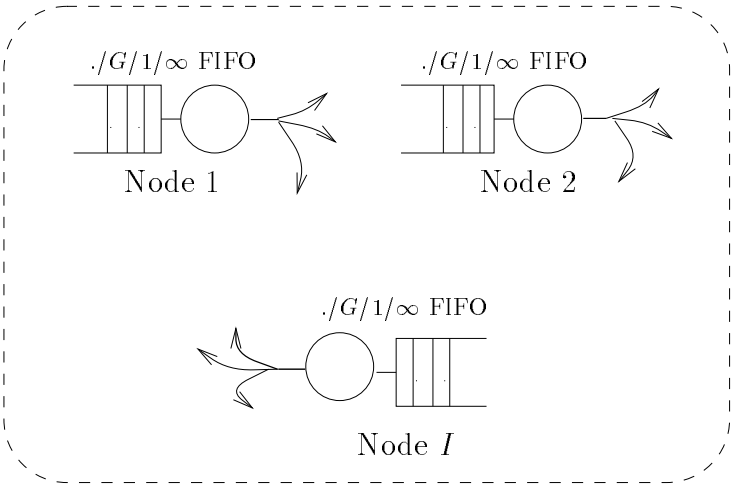


Figure 10.3: A Jackson network.

Customers move from node to node in order to receive some service there. The data are $(2I)$ sequences

$$\{\sigma^i(n), n \in \mathbb{N}\}, \quad \{\nu^i(n), n \in \mathbb{N}\}, \quad i \in \{1, \ldots, I\},$$

where $\sigma^i(n) \in \mathbb{R}^+$ and $\nu^i(n) \in \{1, \ldots, I, I+1\}$.
In the nominal network, the n-th, $n \geqslant 1$, customer to be served by node i after the origin of time requires a service time $\sigma^i(n)$; after completion of its service there, it moves to node $\nu^i(n)$, where $I+1$ is the exit. We say that $\nu^i(n)$ is the n-th routing variable on node i.

We are going to describe the closed (resp. open) Jackson network as a discrete event network (resp. open discrete event network), using the notations of §10.4.

1. **Closed case:** the state at the origin of time is that with all customers in node 1, and service 1 is just starting on node 1. There are no external arrivals and $\nu^i(n) \in \{1, \ldots, I\}$, for all i and n. The total number of customers in the network is then a constant. We take

$$\sigma(n) \stackrel{\text{def}}{=} \sigma^1(n).$$

The *internal daters* $X^{i-}_{[1,\infty]}(n)$ and $X^{i+}_{[1,\infty]}(n)$, $n \geqslant 1$, $i \in \{1, \ldots, I\}$, are the initiation and completion instants of the n-th service on node i. We take

$$X^{-}_{[1,\infty]}(n) \stackrel{\text{def}}{=} X^{1-}_{[1,\infty]}(n),$$

so that $X^{-}_{[1,\infty]}(1) = 0$.

2. **Open case:** the state at the origin of time is that with all queues empty and a customer is just arriving in the network. There is an external arrival point process $\{A_n, n \geqslant 1\}$, with $A_1 = 0$, or equivalently an additional saturated node (numbered 0), which produces customers with inter-arrival times $\sigma^0(n) = A_{n+1} - A_n$, $n \geqslant 1$, regardless of the state of the network. The n-th external arrival is routed to node $\nu^0(n) \in \{1, \ldots, I\}$. We take

$$\sigma(n) \stackrel{\text{def}}{=} \sigma^0(n).$$

We can extend the definition of internal daters, which is the same as above, to $i = 0$ by taking $X^{0-}_{[1,\infty]}(n) = A_n$ and $X^{0+}_{[1,\infty]}(n) = A_n + \sigma(n) = A_{n+1}$. We take

$$X^{-}_{[1,\infty]}(n) \stackrel{\text{def}}{=} X^{0-}_{[1,\infty]}(n)\ ,$$

so that $X^{-}_{[1,\infty]}(1) = 0$.

In both cases, the restrictions $[1, m]$ of the process are obtained by modifying the $\{\sigma(n), n \in \mathbb{N}\}$ sequence in the following way

$$\sigma^i_{[1,m]}(n) = \begin{cases} \sigma^i(n) & \text{for all } n \geqslant 1 \text{ and } i \neq 1 \text{ (resp. } i \neq 0); \\ \sigma^i(n) & \text{for all } 1 \leqslant n \leqslant m \text{ and } i = 1 \text{ (resp. } i = 0); \\ \infty & \text{for all } n > m \text{ and } i = 1 \text{ (resp. } i = 0). \end{cases}$$

The corresponding variables are denoted $X^{-}_{[1,m]}(n), X^{+}_{[1,m]}(n)$. In both cases, the maximal dater is defined as

$$\begin{aligned} X_{[1,m]} \quad = \quad \max \Bigg(& \sup_{i,n} \left\{ X^{i-}_{[1,m]}(n) \text{ s. t. } X^{i-}_{[1,m]}(n) < \infty \right\}, \\ & \sup_{i,n} \left\{ X^{i+}_{[1,m]}(n) \text{ s. t. } X^{i+}_{[1,m]}(n) < \infty \right\} \Bigg), \end{aligned}$$

where the supremum bears on $n \geqslant 1$ and $i \in \{1, \ldots, I\}$ (resp. $i \in \{0, \ldots, I\}$) in the closed (resp. open) case.
The following lemma follows from results proved in [7].

Lemma 10.9.5. *For all $i \in \{1, \ldots, I\}$ and $l \geqslant 1$, there exist finite sets $\mathcal{A}(i,l) \subset \mathbb{N}$, $\mathcal{B}(i,l,p) \subset \mathbb{N}$ where $p \in \mathcal{A}(i,l)$ and $\mathcal{C}(i,l,p,q) \subset \mathbb{N} \times \mathbb{N}$ where $q \in \mathcal{B}(i,l,p)$, which depend on the routing sequences only (not on the service sequences). These sets are such that*

$$\forall m, n \geqslant 1,\ X^{i-}_{[1,m]}(n) = \inf_{l \in \mathcal{A}(i,n)} \max_{p \in \mathcal{B}(i,n,l)} \sum_{(i_q, n_q) \in \mathcal{C}(i,n,l,p)} \sigma^{i_q}_{[1,m]}(n_q). \tag{10.41}$$

A pair (i,n) appears at most once in each set $\mathcal{C}(i,n,l,p)$.

This lemma has to be interpreted as the fact that Jackson networks have a (min,max,+) structure, although a very complicated one. Hence, it should come as no surprise that they enter the monotone separable framework. Let us prove it.

Causality In both cases, the assumption is that $X_{[1,m]}$ is a.s. finite for all m. Note that this implies *causality* as defined in § 10.4.

Lemma 10.9.6. *Causality is satisfied whenever the routing sequences $\{\nu^i(n)\}_{n \in \mathbb{N}}$ are i.i.d. and independent of the service times, and the routing matrix*

$$\mathbb{P} = (p_{ij}), \quad p_{ij} = P(\nu^i(1) = j), \quad i, j \in \{1, \ldots, I\}$$

is without capture in the open case, and irreducible in the closed case.

Proof. The proof is based on the following coupling idea: consider a Kelly network (i.e. a route is attached to a customer, see [96]) where the routes are independent and sampled according to the stopped Markov chain with transition matrix $\mathbb{P}$. By this we mean that in the $[1,m]$-network, the route of the first customer to leave node 1 (resp. 0) is

$$\begin{aligned} &\{N_0 = 1, N_1, \ldots, N_{U_1}\} && \text{in the closed case} \\ &\{N_0 = D, N_1, \ldots, N_{U_{I+1}}\} && \text{in the open case,} \end{aligned}$$

where $\{N_p\}$ is a path of the Markov chain $\mathbb{P}$, U_i is the return time to state i, and D is an independent random variable on $\{1, \ldots, I\}$, with distribution $\pi(i) = P(\nu^0(1) = i)$. The routes of the m first customers to be served at node 1 (resp. to arrive from node 0) are assumed to be independent and identically distributed. In this Kelly network, the routes of these m customers are not affected by the service times (in contrast with what happens in the initial network). Thus, in the closed (resp. open) case, all m customers eventually return to node 1 (resp. leave) provided $\mathbb{P}$ is irreducible (resp. $\mathbb{P}$ is without capture). In addition, such a Kelly network is identical in law to the $[1,m]$ restriction of the original network. So $P(X_{[1,m]} < \infty) = 1$. □

In what follows, we will adopt the assumptions of Lemma 10.9.6 and assume in addition that the service times are integrable.

Monotonicity As an immediate corollary of Lemma 10.9.5, for all fixed routing sequences, for all $m, n \geqslant 1$ and i, the variable $X^{i-}_{[1,m]}(n)$ (and therefore $X^{i+}_{[1,m]}(n)$ as well) is a monotone non-decreasing function of $\{\sigma^j(n),\ j \in [2,\ldots,I], n \geqslant 1,\ \sigma^1(n),\ 1 \leqslant n \leqslant m\}$ (resp. $\{\sigma^j(n),\ j \in [1,\ldots,I], n \geqslant 1,\ \sigma^0(n),\ 1 \leqslant n \leqslant m\}$). This monotonicity extends to the maximal dater as well.

Non-expansiveness Let $j \leqslant I$ and $l \geqslant 1$ be fixed. Consider $\sigma^j(l)$ as a variable and all other service times as constants. Then, it follows from Lemma 10.9.5 that $X^{i-}_{[1,m]}(n)$ is a $(\min, \max)$ function of $\sigma^j(l)$. Thus non-expansiveness as defined in § 10.4 holds.

Separability Let $\varphi^i_{[1,m]} = \sup\{n \geqslant 1 \mid X^{i+}_{[1,m]}(n) < \infty\}, \quad m \geqslant 1$, (the total number of events which ever complete on station i in the $[1,m]$-network). Of course $\varphi^1_{[1,m]} = m$ in the closed case, and $\varphi^0_{[1,m]} = m$ in the open case.
The following two properties hold:

1. For all i and m, $\varphi^i_{[1,m]}$ does not depend on the (finite) values of the variables $\{\sigma^j(n),\ j \in [2,\ldots,I], n \geqslant 1,\ \sigma^1(n),\ 1 \leqslant n \leqslant m\}$ (resp. $\{\sigma^j(n),\ j \in [1,\ldots,I], n \geqslant 1,\ \sigma^0(n),\ 1 \leqslant n \leqslant m\}$) –this follows from Lemma 10.9.5.

2. For all $m \geqslant 1$, the random variables $\{\varphi^i_{[1,m]},\ i \leqslant I\}$ form a stopping time of the sequences $\{\nu^i(n),\ i \leqslant I,\ n \geqslant 1\}$ in the sense that

$$\{\varphi^i_{[1,m]} \leqslant n^i,\ i \leqslant I\} \in \mathcal{F}\{\nu^i(l),\ l \leqslant n^i,\ i \leqslant I\},$$

   where $\mathcal{F}(u)$ denotes the σ-algebra generated by the random variable u.

We are now in a position to complete the definition of $N = \{\sigma(n), M(n),\ n \in \mathbb{N}^*\}$ (see §10.4) for this network, by taking

$$M(n) \stackrel{\text{def}}{=} \{\sigma^i(l),\ \nu^i(l),\ l = \varphi^i_{[1,n-1]} + 1, \ldots, \varphi^i_{[1,n]},\ i \leqslant I\}, \quad n \geqslant 1,$$

with the convention $\varphi^i_{[1,0]} = 0$.
With this definition, the $[m,\infty]$-network, $1 \leqslant m$, is a Jackson network as defined above, but with the driving sequences

$$\begin{aligned} \sigma^i_{[m,\infty]}(n) &= \sigma^i(n + \varphi^i_{[1,m-1]}), \quad n \geqslant 1, \\ \nu^i_{[m,\infty]}(n) &= \nu^i(n + \varphi^i_{[1,m-1]}), \quad n \geqslant 1. \end{aligned}$$

From the i.i.d. assumptions on the sequences $\{\sigma^i(n), \nu^i(n), n \in \mathbb{N}\}$ and the fact that the r. v. $\varphi^i_{[1,m-1]}$ are stopping times, we obtain that the $[m,\infty]$-network is equal in distribution to the original $[1,\infty]$-network. Separability is now clear:

- **Open case:** if $A_{l+1} \geqslant A_1 + X_{[1,l]}$, then from monotonicity, for all i,

$$A_{l+1} \geqslant A_1 + X^{i+}_{[1,l]}(\varphi^i_{[1,l]}) \geqslant A_1 + X^{i+}_{[1,n]}(\varphi^i_{[1,l]}),$$

  and so, the $(l+1)$-st external arrival finds an empty network (we know that if there are l external arrivals and $\varphi^i_{[1,l]}$ departures from node i, then the network is empty). In addition, the next customer to be served on node i is that with index $\varphi^i_{[1,l]} + 1$, $i \leqslant I$. Thus $A_1 + X_{[1,m]} = A_{l+1} + X_{[l+1,m]}$.

- **Closed case:** if $X^+_{[1,m]}(l+1) \geqslant X_{[1,l]}$, then

$$X^+_{[1,m]}(l+1) \geqslant X^{i+}_{[1,l]}(\varphi^i_{[1,l]}) \geqslant X^{i+}_{[1,m]}(\varphi^i_{[1,l]}),$$

  and so, by the same argument as above, when the $(l+1)$-st service ends on node 1, all customers are present in node 1. Separability follows in a way which is similar to that of the previous case.

First order ergodic theorem

Compatibility is immediate from Property 2 of $\{\varphi^i_{[1,m]}\}$. To prove integrability, it is enough to prove that $X_{[1,1]}$ is integrable. This follows from the fact that the stopping times U_1 (resp. U_{I+1}) of $\mathbb{P}$ are integrable and from the assumption that service times are integrable.
Therefore, Theorem 10.4.4 applies and

$$\lim_{m\to\infty} \frac{X_{[1,m]}}{m} = \gamma, \quad \text{a.s.}$$

for some positive and finite constant, both in the open and closed cases. More generally, it can be shown that the above limit implies that there exist finite constants rates γ^i such that

$$\lim \frac{X^{i-}_{[1,m]}}{m} = \gamma^i, \quad \text{a.s.}, \quad i \leqslant I,$$

both in the open and closed cases. For more details on the explicit computation of these rates see [9] and [**?**].

Bibliographie

[1] M. Ajmone-Marsan, G. Balbo, and G. Conte. *Performance Models of Multiprocessor Systems*. MIT Press, Cambridge, Mass., 1986.

[2] V. Anantharam and T. Konstantopoulos. Stationary solutions of stochastic recursions describing discrete event systems. In *Proc. 33rd Conf. on Decision and Control*, volume 2, pages 1481–1486, Lake Buena Vista, FL, 1994.

[3] S. Asmussen. On coupling and weak convergence to stationarity. *Annals of Applied Probability*, 2(3):739–751, 1992.

[4] F. Baccelli. Ergodic theory of stochastic Petri networks. *Annals of Probability*, 20(1):375–396, 1992.

[5] F. Baccelli, A. Borovkov, and J. Mairesse. On large tandem queueing systems. Working paper.

[6] F. Baccelli and P. Brémaud. *Elements of Queueing Theory*. Number 26 in Applications of Mathematics. Springer Verlag, Berlin, 1994.

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

[8] F. Baccelli, G. Cohen, G.J. Olsder, and J.P. Quadrat. *Synchronization and Linearity*. John Wiley & Sons, New York, 1992.

[9] F. Baccelli and S. Foss. Ergodicity of Jackson-type queueing networks. *Queueing Systems*, 17:5–72, 1994.

[10] F. Baccelli and S. Foss. On the saturation rule for the stability of queues. *J. Appl. Prob.*, 32(2):494–507, 1995.

[11] F. Baccelli, S. Foss, and B. Gaujal. Structural, temporal and stochastic properties of unbounded free-choice Petri nets. Technical Report 2411, INRIA, Sophia Antipolis, France, 1994. To appear in IEEE TAC.

[12] F. Baccelli, S. Foss, and J. Mairesse. Jackson networks under stationary and ergodic assumptions. In preparation, 1995.

[13] F. Baccelli and Z. Liu. On the stability condition of a precedence-based queuing discipline. *Adv. Appl. Prob.*, 21:883–898, 1989.

[14] F. Baccelli and Z. Liu. On a class of stochastic recursive equations arising in queueing theory. *Annals of Probability*, 21(1):350–374, 1992.

[15] F. Baccelli and J. Mairesse. Ergodic theory of stochastic operators and discrete event networks. In J. Gunawardena, editor, *Idempotency*. Cambridge Univ. Press, 1995. INRIA Technical Report RR-2641.

[16] N. Bambos. On closed ring queueing networks. *J. Appl. Prob.*, 29:979–995, 1992.

[17] N. Bambos and J.Walrand. Scheduling and stability aspects of a general class of parallel processing systems. *Adv. Appl. Prob.*, 25:176–202, 1993.

[18] N. Bambos and J. Walrand. On stability and performance of parallel processing systems. *J. Ass. for Comp. Machinery*, 38(2):429–452, 1991.

[19] T. Başar and G.J. Olsder. *Dynamic Non-cooperative Game Theory.* Academic Press, 1995. 2nd edition.

[20] C. Berge. *Hypergraphes, Combinatoires des ensembles finis.* Gauthier-Villars, Paris, 1987.

[21] J. Berstel and C. Reutenauer. *Rational Series and their Languages.* Springer Verlag, 1988.

[22] A. Borovkov. *Asymptotic Methods in Queueing Theory.* John Wiley & Sons, New York, 1984.

[23] A. Borovkov. Limit theorems for queueing networks. I. *Theory Prob. Appl.*, 31:413–427, 1986.

[24] A. Borovkov. Limit theorems for queueing networks. II. *Theory Prob. Appl.*, 32:257–272, 1988.

[25] A. Borovkov and S. Foss. Stochastically recursive sequences and their generalizations. *Siberian Adv. in Math.*, 2:16–81, 1992.

[26] A. Borovkov and S. Foss. Two ergodicity criteria for stochastically recursive sequences. *Acta Applicandae Mathematicae*, 34:125–134, 1994.

[27] P. Bougerol and J. Lacroix. *Products of Random Matrices with Applications to Schrödinger Operators.* Progress in Probability and Statistics. Birkäuser, 1985.

[28] H. Braker. *Algorithms and Applications in Timed Discrete Event Systems.* PhD thesis, Delft Univ. of Technology, 1993.

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

[30] G. W. Brams. *Réseaux de Petri, Théorie et pratique, Tome 1.* Masson, Paris, 1983.

[31] A. Brandt. On stationary waiting time and limiting behaviour of queues with many servers I: the general G/G/m/∞ case. *Elektron. Inform. Kybern.*, 21:47–64, 1985.

[32] A. Brandt, P. Franken, and B. Lisek. *Stationary Stochastic Models.* Prob. and Math. Stat. Wiley, New York, 1990.

[33] M. Brilman and J.M. Vincent. Synchronisation by resources sharing : a performance analysis. Technical report, MAI-IMAG, Grenoble, France, 1995.

[34] M. Canales and B. Gaujal. Marking optimization and parallelism in marked graphs. Technical Report RR-2049, INRIA, Sophia Antipolis, France, 1993.

[35] P. Cartier and D. Foata. *Problèmes combinatoires de commutation et réarrangements.* Number 85 in Lecture Notes in Mathematics. Springer Verlag, 1969.

[36] C. Cérin and A. Petit. Speedup of recognizable trace languages. In *Proc. MFCS 93*, number 711 in Lect. Notes Comput. Sci. Springer, 1993.

[37] C. S. Chang. On the exponentiality of stochastic linear systems under the Max-Plus algebra. Working paper, 1995.

[38] D. Cheng and D. Yao. Tandem queues with general blocking: A unified model and comparison results. *DEDS: Th. and Appl.*, 2:207–234, 1993.

[39] N. Chernova and S. Foss. Ergodic properties of polling systems. Technical report, Novosibirsk University, 1994. Submitted to Problems of Information Transmission.

[40] W. Chou and R. Duffin. An additive eigenvalue problem of physics related to linear programming. *Adv. in Appl. Math.*, 8:486–498, 1987.

[41] W. Chou and R. Griffiths. Ground states of one-dimensional systems using effective potentials. *Physical Review B*, 34(9):6219–6234, 1986.

[42] P. Chretienne. *Les Réseaux de Petri Temporisés*. PhD thesis, Université Paris VI, Paris, 1983.

[43] G. Cohen, D. Dubois, J.P. Quadrat, and M. Viot. Analyse du comportement périodique des systèmes de production par la théorie des dioïdes. Technical Report 191, INRIA, 1983.

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

[45] G. Cohen, P. Moller, J.P. Quadrat, and M. Viot. Linear system theory of discrete-event systems. In *23rd IEEE CDC, Las Vegas, Nevada*, volume 1, pages 539–544, 1984.

[46] J.E. Cohen. Subadditivity, generalized product of random matrices and operations research. *SIAM Review*, 30(1):69–86, 1988.

[47] J.H. Conway. *Regular algebra and finite machines*. Chapman and Hall, 1971.

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

[49] R. Cuninghame-Green. Describing industrial processes with interference and approximating their steady-state behaviour. *Oper. Res. Quat.*, 13(1):95–100, 1962.

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

[51] H. Daduna. On the stability of queuing systems under precedence restrictions for the service of customers. *Queueing Systems*, 17:73–88, 1994.

[52] A. Darte, L. Khachiyan, and Y. Robert. Linear scheduling is nearly optimal. *Parallel Processing Letters*, 1(2):73–81, 1991.

[53] A. Darte and F. Vivien. Automatic parallelization based on multi-dimensional scheduling. Technical Report 24, LIP, ENS Lyon, September 1994.

[54] M. de Groot. Reaching a consensus. *J. Amer. Stat. Assoc.*, 69:118–121, 1974.

[55] E. Denardo. Period of connected networks. *Math. Oper. Res.*, 2:20–24, 1977.

[56] V. Diekert. *Combinatorics on traces*. Number 454 in LNCS. Springer Verlag, 1990.

[57] E. Dijkstra. Cooperating sequential processes. In F. Genvys, editor, *Programming languages*, pages 43–112. Academic Press, New York, 1968.

[58] P. Dudnikov. Endomorphisms of the semimodule of bounded functions. In Maslov and Samborskiĭ, editors, *Idempotent analysis*, volume 13 of *Adv. in Sov. Math.* AMS, 1992.

[59] V. Dumas. A multiclass network with non-linear, non-convex, non-monotonic stability conditions. Technical Report RR-2487, INRIA, Rocquencourt, France, 1995.

[60] S. Eilenberg. *Automata, languages and machines*, volume A. Academic Press, New York, 1974.

[61] M. Ekhaus and L. Gray. Convergence to equilibrium and a strong law for the motion of restricted interfaces. Working paper.

[62] P. Erdös, P. Goodman, and L. Pósa. The representation of a graph by set intersections. *Canad. J. Math.*, 18:106–112, 1966.

[63] M. Fliess. *Sur certaines familles de séries formelles.* Thèse de doctorat d'etat, Université Paris VII, 1972.

[64] R. Floyd. Algorithm 97, shortest path. *Comm. ACM*, 5:345, 1962.

[65] L.R. Ford and D.R. Fulkerson. *Flows in Networks.* Princeton University Press, 1962.

[66] H. Furstenberg and H. Kesten. Products of random matrices. *Ann. Math. Statist.*, 31:457–469, 1960.

[67] S. Gaubert. *Théorie des systèmes linéaires dans les dioïdes.* PhD thesis, École des Mines, Paris, 1992.

[68] S. Gaubert. *Introduction aux systèmes dynamiques à événements discrets.* Polycopié de cours ENSTA, 1993.

[69] S. Gaubert. Performance evaluation of timed automata. Technical Report 1922, INRIA, 1993. To appear in IEEE TAC.

[70] S. Gaubert. On semigroups of matrices in the (max, +) algebra. Technical Report 2172, INRIA, 1994.

[71] S. Gaubert and J. Mairesse. Scheduling and simulation of timed conservative Petri nets using (max,+) automata. In preparation, 1995.

[72] S. Gaubert and J. Mairesse. Task resource models and (max,+) automata. In J. Gunawardena, editor, *Idempotency.* Cambridge University Press, 1995.

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

[74] B. Gaujal, A. Jean-Marie, and J. Mairesse. Minimal representation of uniform recurrence equations. Technical Report RR-2568, INRIA, Sophia-Antipolis, France, 1995.

[75] P. Glasserman and D. Yao. *Monotone Structure in Discrete-Event Systems.* John Wiley & Sons, 1994.

[76] G. Golub and C. Van Loan. *Matrix Computations.* North Oxford, Oxford, 1983.

[77] M. Gondran and M. Minoux. Valeurs propres et vecteurs propres dans les dioïdes et leur interprétation en théorie des graphes. *EDF, Bulletin de la Direction des Etudes et Recherches, Serie C, Mathématiques Informatique*, 2:25–41, 1977.

[78] M. Gondran and M. Minoux. *Graphes et algorithmes.* Eyrolles, Paris, 1979. Engl. transl. *Graphs and Algorithms*, Wiley, 1986.

[79] W. Gordon and G. Newell. Closed queuing systems with exponential servers. *Oper. Res.*, 15:254–265, 1967.

[80] R. Griffiths. Frenkel-Kontorova models of commensurate-incommensurate phase transitions. In H. van Beijeren, editor, *Fundamental problems in statistical mechanics VII.* Elsevier Science Publishers, 1990.

[81] J. Gunawardena. Min-max functions. Technical Report STAN-CS-93-1474, Stanford University, 1993. To appear in Discrete Event Dynamic Systems.

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

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

[84] C. Hanen and A. Munier. Cyclic scheduling on parallel processors: an overview. In P. Chretienne, E. Coffman, J. Lenstra, and Z. Liu, editors, *Scheduling Theory and Its Applications*. John Wiley, 1995.

[85] C. Hanen and A. Munier. A study of the cyclic scheduling problem on parallel processors. *Discrete Appl. Math.*, 57:167–192, 1995.

[86] F. Harary. *Graph Theory*. Addison-Wesley, 1969.

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

[88] Y.C. Ho, editor. *Special Issue on Discrete Event Dynamic Systems*, volume 77. Proceedings of the IEEE, 1989.

[89] A. Hoffman and J. Kruskal. Integral boundary points of convex polyhedra. *Ann. of Math. Studies, Princeton*, 38:223–246, 1956.

[90] J.R. Jackson. Jobshop-like queueing systems. *Manag. Science*, 10:131–142, 1963.

[91] A. Jean-Marie and G.J. Olsder. Analysis of stochastic min-max systems : results and conjectures. Technical Report 93-94, Faculty Technical Math. and Informatics, Delft University of Technology, 1993.

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

[93] R. Karp, R. Miller, and S. Winograd. The organization of computations for uniform recurrence equations. *J. Ass. Comp. Mach.*, 14(3):563–590, 1967.

[94] H. Kaspi and A. Mandelbaum. Regenerative closed queueing networks. *Stoch. and Stoch. Reports*, 39:239–258, 1992.

[95] H. Kaspi and A. Mandelbaum. On Harris recurrence in continuous time. *Math. Oper. Research (to appear)*, 1994.

[96] F.P. Kelly. *Reversibility and Stochastic Networks*. Wiley, New-York, 1979.

[97] J. Kingman. Subadditive ergodic theory. *Annals of Probability*, 1:883–909, 1973.

[98] V. Kolokoltsov. On linear, additive and homogeneous operators in idempotent analysis. In V. Maslov and S. Samborskiĭ, editors, *Idempotent Analysis*, volume 13 of *Adv. in Sov. Math.* AMS, 1992.

[99] T.E Lee. Stable earliest starting schedules for periodic job shops: a linear approach. In G. Cohen and J.P. Quadrat, editors, *11th Int. Conf. on Analysis and Optim. of Systems*, number 199 in Lecture Notes in Control and Info. Sciences, pages 290–296, Sophia-Antipolis, France, 1994. Springer Verlag.

[100] C. Leiserson and J. Saxe. Retiming synchronous circuitry. *Algorithmica*, 6:5–35, 1991.

[101] T. Liggett. *Interacting Particle Systems.* Springer Verlag, Berlin, 1985.

[102] T. Lindvall. *Lectures on the Coupling Method.* John Wiley & Sons, New-York, 1992.

[103] R. Loynes. The stability of a queue with non-independent interarrival and service times. *Proc. Camb. Philos. Soc.*, 58:497–520, 1962.

[104] J. Mairesse. A graphical representation of matrices in the (max,+) algebra. Technical Report RR-2078, INRIA, Sophia Antipolis, France, 1993. Available via anonymous ftp at ftp.inria.fr in INRIA/publication/RR/RR-2078.ps.gz.

[105] J. Mairesse. Products of irreducible random matrices in the (max,+) algebra. Technical Report RR-1939, INRIA, Sophia Antipolis, France, 1993. To appear in *Adv. Applied Prob.*

[106] J. Mairesse. A graphical approach of the spectral theory in the (max,+) algebra. *IEEE Trans. Automatic Control*, 40(10), October 1995.

[107] V. Maslov and S. Samborskiĭ, editors. *Idempotent analysis*, volume 13 of *Adv. in Sov. Math.* AMS, 1992.

[108] L. McKenzie. Turnpike theory. *Econometrica*, 44(5):841–865, 1976.

[109] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, Berlin, 1993.

[110] W. Miller. The maximum order of an element of a finite symmetric group. *American Math. Monthly*, 94:497–506, 1987.

[111] E. Moore. The shortest path through a maze. In *Proc. Int. Symp. on Theory of Switching, Part II*, pages 285–292, Austin, Texas, 1957.

[112] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989.

[113] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson, Paris, 1970.

[114] G.J. Olsder. Eigenvalues of dynamic min-max systems. *JDEDS*, 1:177–207, 1991.

[115] G.J. Olsder, J. Resing, R. de Vries, M. Keane, and G. Hooghiemstra. Discrete event systems with stochastic processing times. *IEEE Trans. on Automatic Control*, 35(3):299–302, 1990.

[116] K. Parhi and D. Messerschmidt. Static rate-optimal scheduling of iterative Data-flow programs via optimum unfolding. *IEEE Trans. on Computers*, 1991.

[117] C. A. Petri. *Kommunikation mit Automaten.* PhD thesis, Institut fur Instrumentelle Mathematik, Bonn, 1962. English translation "Communication with Automata", Griffiss Air Foce Base, New York, Tech. Rep. RADC TR-65-377, vol. 1, suppl. 1, 1966.

[118] J.E. Pin. *Varieties of Formal Languages.* Plenum Pub. Corp., New-York, 1986.

[119] N. Pippenger. Pebbling. Research report RC 8258 (#35937), IBM Research Division, 1980.

[120] P. Ramadge and W. Wonham. The control of discrete event systems. *Proc. IEEE*, 77:81–98, 1989.

[121] R. Reiter. Scheduling parallel computations. *JACM*, 15(4):590–599, 1968.

[122] J. Resing, R. de Vries, G. Hooghiemstra, M. Keane, and G.J. Olsder. Asymptotic behavior of random discrete event systems. *Stoch. Proc. and Applications*, 36:195–216, 1990.

[123] I.V. Romanovskiĭ. Optimization and stationary control of discrete deterministic process in dynamic programming. *Cybernetics*, 3:66–78, 1967.

[124] S. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, New York, 1983.

[125] A. Rybko and A. Stolyar. Ergodicity of stochastic processes describing the operations of open queueing networks. *Problemy Peredachi Informatsii*, 28:3–26, 1992.

[126] N. Saheb. Concurrency measure in commutation monoids. *Discrete Applied Mathematics*, 24:223–236, 1989.

[127] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer series in statistics. Springer Verlag, Berlin, 1981.

[128] R. Sethi. Complete register allocation problems. *SIAM J. Comp.*, 4(3):226–248, 1975.

[129] N. Sheney and R. Rudell. Efficient implementation of retiming. working paper, 1994.

[130] R. Sine. A non-linear Perron-Frobenius theorem. *Proc. AMS*, 109(2):331–336, 1990.

[131] S. Sriram and E.A. Lee. Statically scheduling communication resources in multiprocessor DSP architectures. In *28-th Conference on Signals, Systems and Computers, Asilomar*, 1994.

[132] R. Tarjan. Depth-first search and linear graph algorithm. *SIAM J. Computing*, 1:146–160, 1972.

[133] F. Tronel. Domaines d'attraction des opérateurs (max,+) linéaires. Technical report, ENS Lyon, 1995. Rapport de stage, direction J. Mairesse.

[134] G.X. Viennot. Heaps of pieces, I: Basic definitions and combinatorial lemmas. In Labelle and Leroux, editors, *Combinatoire Énumérative*, number 1234 in Lect. Notes in Math., pages 321–350. Springer, 1986.

[135] G.X. Viennot. Trees. In M. Lothaire, editor, *Mots*, pages 265–297, Paris, 1990. Hermès.

[136] J.M. Vincent. Some ergodic results on stochastic iterative DEDS. Technical Report 4, Apache, IMAG, Grenoble, 1994. To appear in JDEDS.

[137] J.M. Vincent. Dynamique de systèmes de réservation. working paper, 1995.

[138] D. Weller. *Hilbert's metric, part metric and self-mappings of a cone*. PhD thesis, University of Bremen, Germany, 1987.

[139] C. Wende, Q. Xiangdong, and D. Shuhui. The eigen-problem and period analysis of the discrete event systems. *Systems Sci. and Mathematical Sci.*, 3(3), 1990.

[140] D.J. White. *Markov Decision Processes*. Wiley, New York, 1993.

[141] W. Willinger. Traffic modeling for high-speed networks: theory versus practice. In F. Kelly and R. Williams, editors, *Stochastic Networks*, volume 71 of *IMA*. Springer-Verlag, 1995.

[142] S. Yakovenko and L. Kontorer. Nonlinear semigroups and infinite horizon optimization. In V. Maslov and S. Samborskiĭ, editors, *Idempotent analysis*, volume 13 of *Adv. in Sov. Math.* AMS, 1992.

Index

$B(A,.)$ (ball), 139
$G/G/1/\infty$, 207
$\mathbb{PR}^k$, 78
$\mathcal{B}(A,.)$ (projective ball), 154
$\cdot/\cdot/1/\infty$, 23
ε, 73
$|.|_{\mathcal{P}}$, 124, 212
$\odot$, 221
$\oplus$, 15, 73
$\otimes$, 15, 73
$\pi(.)$, 78
$\mathbb{R}_{\max}$, 15, 73
e, 73
θ, 197
$d(.,.)$, 79, 124
$\mathtt{D}(.)$, 125

algèbre
    (max,+), 15
    de chemin, 16
algebra
    (max,+), 73, 123
alphabet, 27, 178
assembly line, 74–75, 95–96
atelier flexible, 21–23
automate, 26–31
automaton, 181
    (max,+), 179

basis (change of), 81
Bellman (équation de), 32
birth and death process, 145
blocage, 23–26
Borovkov theory, 129
Brandt maximal solution, 214
Burnside problem, 96

Cartier-Foata normal form, 178
causality, 203
characteristic series, 187
coût, 31
configuration, 43
convergence
    étroite, 9
    coupling, 129
    strong coupling, 129
    total variation, 9, 129, 136
    variation, 9
    weak, 9, 131, 136, 149
coupling
    ε-, 131
    η-, 131, 136
critical circuit, 76
cut, 48
    compatible, 50
    consecutive, 49
    span of a, 52
cycle time, 75, 102, 119, 126
cyclicity, 76, 77
    maximal value of, 113

dater, 180
degree of parallelism, 41
dependence alphabet, 178
dependence relation, 178
digital circuit, 60–64
dining philosopher, 179
dioid, 123
discrete event network, 202
    monotone–separable, 203
    open, 205
dynamic programming, 222

eigenvalue, 74
eigenvector, 74
    extremal, 77
ergodic lemma, 197
event graph, 66, 171
    FIFO, 67
    stochastic, 172
execution, 43
    asynchronous, 44
    synchronous, 45

synchronous regular, 45

FIFO, 67
file d'attente, 23–26
firing count vector, 171
first (second) order variable, 118, 121, 183, 196
fixed point, 214
generalized, 200
fixed structure, 124
flow, 48
free monoid, 178
Frenkel-Kontorova (modèle de), 33
functional element or gate, 60

Gagny's triangle, 190
Gantt charts, 103
Goldstein (théorème de), 9
graph
2-section of a, 192
bipartite, 191
clique of a, 191
communication, 123
complete, 191
critical, 76
dependence, 38
event, 66
of a matrix, 76
reduced, 39
simple, 191
strongly connected, 76
task, 120
graphe d'événements, 18
vivacité d'un, 18
graphe des marquages, 30

Hadamard product, 187
Harris ergodicity, 130
heap of pieces, 182
Hilbert projective metric, 124
homogeneity, 198
sub–, 204
hypergraph, 182

idle time, 121
internal monotonicity, 209

Jackson network, 120, 223
cyclic, 120, 127, 137, 140, 144
jeton, 17
job-shop, 21–23

Kelly network, 225
Kendall (notation de), 23
Kingman theorem, 197

langage, 27
reconnu, 27
language
algebraic, 187
context-free, 187
rational, 187
recognizable, 179
latency, 102
level (of a dependence graph), 38
Lindley's equation, 207
linear combination, 78, 80
Loynes scheme, 147, 207
Loynes variable, 208
Lyapunov exponent, 126, 132
optimal, 186
worst, 186

mécanique statistique, 33–35
makespan, 65, 178
marching soldier model, 221
marking, 66
Markov chain, 130
marquage, 17
Maslov (intégrale de), 35
mathematical trivia, 168
matrix
aperiodic, 76
basic, 82
Boolean, 128
change of basis, 81
discretization, 150
fixed structure, 124
incidence, 171
irreducible, 76
markovian, 219
normalized, 76, 126
permutation, 80
rank of a, 127
semigroup of, 96
skeleton, 150
stationary, 77
max,+ algebra, 73, 123
merci (beaucoup), 1
minimal value, 211
mise en abîme polyphonique, 10
monoid, 178

monotonicity, 198, 203

non-expansiveness, 198, 203

operator, 198
  (+,×), 220
  (max,+), 220
  (max,+,×), 221
  (min,max,+), 221
  (min,max,+,×), 219
  bounded, 125
optimisation dynamique, 31–33

PAPS, 21
path, 48
pattern, 136, 139
  asymptotic, 139
pebble game, 42–46
  configuration, 43
  execution, 43
periodic regime, 74
Petri
  net, 66
  réseau de, 17
  réseau de – avec choix, 29–31
place, 17, 66
Polish space, 129
polonais (espace), 9
power algorithm, 113
precedence constraints model, 222
primitive, 128
programmation dynamique, 31–33
projective
  boundedness, 200
  diameter, 125
  distance, 79, 124
  size, 82
  space, 78, 124
pseudo-periodic model, 138

railway network, 190
rank, 127
reachability, 171
recognizable, 179
register, 60
renovating event, 129
retiming, 53
  legal, 63
reward, 222
ring network, 179, 185

s.c.s., 111, 133
schedule, 42, 102, 178
  fully-static, 104
  linear, 42
  optimal, 186
  self-timed, 104
  worst, 186
scheduling, 101
  cyclic, 101
scs-cyc, 77
  scs1-cyc2, 87–91
  scs1-cyc3, 84–87
  scs2-cyc1, 87–91
  scs2-cyc2, 91–92
  scs3-cyc1, 84–87
scs1-cyc1, 83, 127
section, 48
  compatible, 50
semigroup, 96, 128
  primitive, 128
separability, 203
  strong, 208
shift, 197
simulation, 67
span, 52
spectral theory, 75
SRS, 129
stabilité, 9
stability, 121, 136, 208
stationary regime, 77, 130, 136
  weak, 130
stochastic recursive sequence, 129
subadditivity, 197
subdiagonal, 86
synchronic distance, 171
systèmes à événements discrets, 7

task graph, 120
task resource system, 177
Tetris model, 188
throughput, 75, 119, 121
tightness, 130
token, 66
totally unimodular, 193
trace monoid, 178
transformation
  backward, 70
  forward, 70
transient regime, 77, 92–94
transition, 17, 66
  firing of a, 66

habilitation d'une, 17
tir d'une, 17
transition de phase, 35
turnpike theory, 32

Uniform Recurrence Equation (URE), 38
constructive, 38
delay, 37
recycled, 40
utilité, 31

valeur propre, 16
vecteur propre, 16

word, 178
workload, 121, 207