Contrôle optimal dans les réseaux et mots équilibrés

Bruno Gaujal Alain Jean-Marie

ID-IMAG, INRIA Grenoble et INRIA/LIRMM Montpellier

http://www-id.imag.fr/Laboratoire/Membres/Gaujal_Bruno/perso.html
http://www.lirmm.fr/~ajm

École Jeunes Chercheurs en Algorithmique et Calcul Formel, Montpellier
6 avril 2005

Plan du cours

- Partie I : Fondements théoriques
    1. Mots équilibrés
    2. Multimodularité
    3. Le théorème de minimisation
- Partie II : Applications aux systèmes à événements discrets
    1. Contrôle des systèmes à événements discrets (rejet, routage, polling dans les réseaux de télécommunication).
    2. Problèmes de calculs
    3. Plusieurs extensions (problèmes multidimensionnels, ordres de régularité, contrôle en boucle fermée, multimodularité en temps et en espace,...).

Mots équilibrés

Soit l'alphabet $\mathcal{A} = \{0, 1\}$. $\mathcal{A}^{\mathbb{N}}$ ($\mathcal{A}^{\mathbb{Z}}$) est l'ensemble des mots (ou suites) binaires (bi-infinies). Le support du mot u est $\mathcal{S}_1 = \{i \mid u_i = 1\}$.

Définition

La suite u est équilibrée si pour tout ℓ et pour toute paire W et W' de facteurs de u de taille ℓ,

$$-1 \leqslant |W|_1 - |W'|_1 \leqslant 1.$$

Un mot de Sturm est une suite équilibrée non périodique.

Lemme

Si u est équilibrée alors

- $-1 \leqslant |W|_0 - |W'|_0 \leqslant 1.$
- $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} u_n$ *existe (densité de u).*

Mots crochets sur $\mathbb{Z}$

Théorème

[Morse Hedlund] *Une suite $u \in \{0,1\}^{\mathbb{Z}}$ est équilibrée de densité $d = 1/\beta$ ssi le support $\mathcal{S}_1$ satisfait un des cas suivants :*

(a) *(cas irrationnel) d est irrationnel et il existe $\phi \in \mathbb{R}$ tel que*

$$\mathcal{S}_1 = \{\lfloor i\beta + \phi \rfloor\}_{i\in\mathbb{Z}} \ \text{ ou } \ \mathcal{S}_1 = \{\lceil i\beta + \phi \rceil\}_{i\in\mathbb{Z}}.$$

(b) *(cas périodique) $d \in \mathbb{Q}$ et il existe $\phi \in \mathbb{Q}$ tel que*

$$\mathcal{S}_1 = \{\lfloor i\beta + \phi \rfloor\}_{i\in\mathbb{Z}}.$$

(c) *(torsion) $d \in \mathbb{Q}$ $(d = k/n,\ k, n \in \mathbb{N})$ et il existe $m \in \mathbb{Z}$ tel que*

$$\mathcal{S}_1 = \{\lfloor in/k + m \rfloor\}_{i<k} \cup \{\lfloor in/k - 1/k + m \rfloor\}_{i>0}$$

ou

$$\mathcal{S}_1 = \{\lfloor in/k + m \rfloor\}_{i>0} \cup \{\lfloor in/k - 1/k + m \rfloor\}_{i<k}$$

Mots crochets sur $\mathbb{N}$

Théorème

[Morse Hedlund] Une suite $u \in \{0,1\}^{\mathbb{N}}$ est équilibrée de densité $d = 1/\beta$ ssi le support $\mathcal{S}_1$ satisfait un des cas suivants

(a) *(cas irrationnel) d est irrationnel et il existe $\phi \in \mathbb{R}$ tel que*

$$\mathcal{S}_1 = \{\lfloor i\beta + \theta \rfloor\}_{i \in \mathbb{N}} \ \ \textit{or} \ \ \mathcal{S}_1 = \{\lceil i\beta + \theta \rceil\}_{i \in \mathbb{N}}.$$

(b) *(cas périodique) $d \in \mathbb{Q}$ et il existe $\phi \in \mathbb{Q}$ tel que*

$$\mathcal{S}_1 = \{\lfloor i\beta + \theta \rfloor\}_{i \in \mathbb{N}}.$$

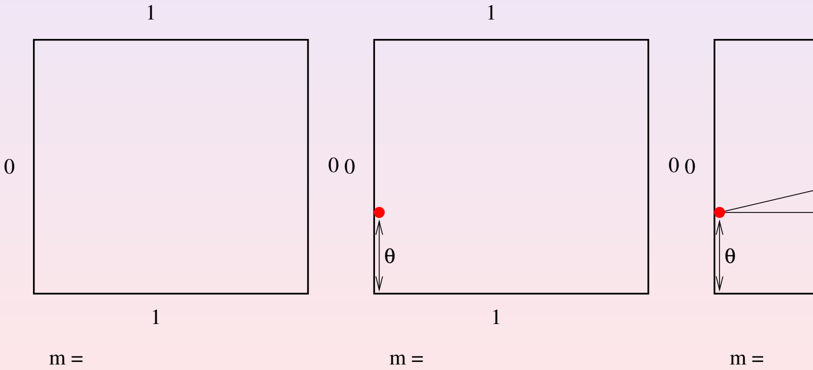
Corollaire

Les suites équilibrées sont ultimement des suites crochets :
Pour n assez grand,

$$m_{\alpha,\phi}(n) = \lceil n\alpha + \phi \rceil - \lceil (n-1)\alpha + \phi \rceil.$$

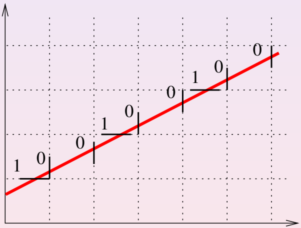
Construction de mots équilibrés

On peut construire les suites crochets comme des suites billard. (avec $s = d/1 - d$ et $\theta = \phi$)



Construction de mots équilibrés (II)

Construction comme droite discrète (suites de Beatty) ou en dépliant la trajectoire du billard.



Construction de mots équilibrés (III)

Construction par morphismes.
Un morphisme Sturmien est un morphisme sur les mots infinis qui transforme un mot de Sturm en un mot de Sturm.

Théorème

[Mignosi, Séébold] Les morphismes Sturmiens sont générés par les trois morphismes Sturmiens suivants

$$E : \left\{ \begin{array}{ccc} 0 & \rightarrow & 1 \\ 1 & \rightarrow & 0 \end{array} \right. \qquad \varphi : \left\{ \begin{array}{ccc} 0 & \rightarrow & 01 \\ 1 & \rightarrow & 0 \end{array} \right. \qquad \tilde{\varphi} : \left\{ \begin{array}{ccc} 0 & \rightarrow & 10 \\ 1 & \rightarrow & 0 \end{array} \right.$$

Construction de mots équilibrés (IV)

Construction par les fractions continues.

On décompose la densité α en fraction continue :

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}.$$

$s_{-1} = 1, s_0 = 0, s_n = s_{n-1}^{a_n} s_{n-2}$: "suite des mots standards" associée à la suite $a_1, \cdots, a_n, \cdots$.

Théorème

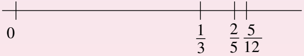
[Stolarsky, 1976] La suite des mots standards est une suite de préfixes du mot équilibré caractéristique, c_α (mot crochet avec $\phi = 1/\alpha$) et $\lim_{n\to\infty} s_n = c_\alpha$.

Fractions continues inférieures

Soit $0 < \alpha < 1$, $\alpha = \langle l_1, l_2, \ldots, l_n, \ldots \rangle = \langle l_1, l_2, \ldots, l_n - \alpha_n \rangle$.

$$\alpha = \cfrac{1}{l_1 - \cfrac{1}{l_2 - \cfrac{1}{l_3 - \alpha_3}}}, \qquad \text{avec } l_{n+1} = \lceil \alpha_n^{-1} \rceil.$$

Par exemple, la décomposition en fraction continue inférieure de $\frac{5}{12}$ est $\langle 3, 2, 3 \rangle$.



Décomposition en facteurs (x-y)

Soit $(x_n)_{n\geqslant 0}$ et $(y_n)_{n\geqslant 0}$ 2 suites binaires telles que

$$x_0 = 1, \qquad x_n = x_{n-1}y_{n-1}^{l_n-2},$$
$$y_0 = 0, \qquad y_n = x_{n-1}y_{n-1}^{l_n-1}.$$

Théorème

[Gaujal, Hyon] Les x_i et les y_i sont des facteurs du mot crochet $m_{\alpha,\phi}$ pour tout ϕ. De plus $y_n \to m_{\alpha,0}$ (noté m_α dans la suite).

Pour $\alpha = 5/12 = \langle 3, 2, 3\rangle$.

$x_1 = 10 \qquad x_2 = x_1 = 10 \qquad x_3 = x_2y_2 = 1010100,$

$y_1 = 100 \qquad y_2 = x_1y_1 = 10100 \qquad y_3 = x_2y_2y_2 = 101010010100.$

$$m_{5/12} = \overbrace{\underbrace{\overbrace{10}^{x_1}}_{x_2}\underbrace{\overbrace{10}^{x_1}\overbrace{100}^{y_1}}_{y_2}\underbrace{\overbrace{10}^{x_1}\overbrace{100}^{y_1}}_{y_2}}^{y_3}.$$

Mots équilibrés à plusieurs lettres

Si l'alphabet $\mathcal{A}$ contient $k > 2$ lettres, alors un mot est équilibré s'il est équilibré pour toutes ses lettres.
Contrairement au cas binaire, il n'existe pas toujours de mots équilibrés de densités $(\alpha_1, \ldots, \alpha_k)$ dès que $\alpha_1 + \cdots + \alpha_k = 1$.
Pour les mots non périodiques, il existe une caractérisation facile des mots équilibrés.

Théorème

[Graham][11] $u \in \mathcal{A}^{\mathbb{Z}}$ est équilibré et non ultimement périodique ssi il existe une partition de $\mathcal{A}$ en $\mathcal{A}_1$ et $\mathcal{A}_2$ telle que $\mathbf{1}_{\mathcal{A}_1}(u)$ est équilibrée et $\Pi_{\mathcal{A}_1}(u)$, $\Pi_{\mathcal{A}_2}(u)$ sont équilibrés avec des densités inverses d'entiers.

Mots équilibrés à plusieurs lettres(II)

Pour les mots périodiques, la situation est plus complexe. On ne sait pas aujourd'hui les caractériser tous.

Conjecture

[Conjecture de Fraenkel] Une suite équilibrée sur $k > 2$ lettres existe, de densités $(\alpha_1, \ldots, \alpha_k)$ toutes distinctes, ssi $\alpha_i = 2^{i-1}/(2^k - 1)$.

Exercices

1. (*) Montrer que s'il existe un mot équilibré sur n lettres de densités $(1/n_1, \cdots, 1/n_k)$, $n_1 \geqslant \ldots \geqslant n_k \in \mathbb{N}$, alors $n_1 = n_2$.
2. Trouver toutes les densités possibles de mots équilibrés sur 3 lettres.

Multimodularité

Dans $\mathbb{R}^n$, $\mathcal{F}$: base multimodulaire : $n+1$ vecteurs de rang n qui somment à 0 : $s_0 + \cdots + s_n = 0$.

$\mathcal{D}_d$ décalages à droite :

$$\begin{array}{lccccc} s_0 = & +1 & 0 & \cdots & 0 & 0 \\ s_1 = & -1 & +1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ s_n = & 0 & 0 & \cdots & 0 & -1 \end{array}$$

M : ensemble des points de coordonnées entières sur $\mathcal{F}$.

Définition

[Hajek][12] Une fonction $f : M \to \mathbb{R}$ est multimodulaire sur $\mathcal{F}$ si

$$\forall x \in M, \quad i \neq j, f(x+s_i) + f(x+s_j) \geqslant f(x+s_i+s_j) + f(x).$$

Multimodularité (II)

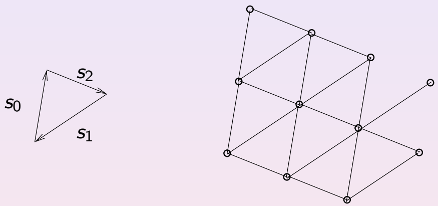


FIG.: Exemple dans $\mathbb{R}^2$.

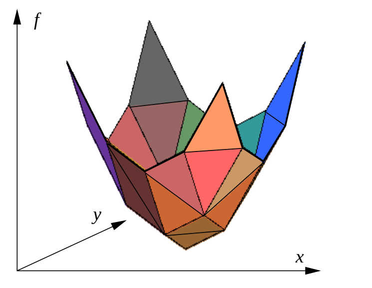
f est définie sur le treillis M des points entiers dans la base $\mathcal{F}$.
$\tilde{f}$ est l'interpolation linéaire de f sur les simplexes.

Multimodularité (III)

Par des arguments de dimensions des faces des simplexes de M,

Théorème

[Altman, Hordijk, G.]
f *est multimodulaire* $\Leftrightarrow$ $\tilde{f}$ *est convexe.*



Remarques

La notion de multimodularité est liée à la base choisie.
Exemple :
max : $\mathbb{Z}^2 \to \mathbb{R}$ est multimodulaire avec $\mathcal{F} = \{(-1,-1),(0,1),(1,0)\}$
pas multimodulaire avec $\mathcal{F} = \{(-1,+1),(0,-1),(1,0)\}$.

Exercices

1. Montrer que si f est multimodulaire pour la base $\mathcal{F}$, alors f est multimodulaire pour la base $-\mathcal{F}$.
2. Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe. Est-ce qu'il existe une base $\mathcal{F}$ telle que le treillis M engendré par $\mathcal{F}$ a tous ses points dans $\mathbb{Z}^n$ et $f : M \to \mathbb{R}$ soit multimodulaire ?
3. (**) Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction convexe. Est-ce qu'il existe une base $\mathcal{F}$ telle que $f : M_{\mathcal{F}} \to \mathbb{R}$ soit multimodulaire ?

Mots équilibrés et décalages à droite

$\mathcal{D}_d$ base des décalages à droite de $\mathbb{R}^n$. Le point $(\alpha, \cdots, \alpha)$ appartient au simplexe avec pour extrémités $p(\alpha, \theta)$, $0 \leqslant \theta \leqslant 1$ (préfixes de mots crochets) :

$$p(\alpha, \theta) = \Big(\lfloor 2\alpha + \theta \rfloor - \lfloor \alpha + \theta \rfloor, \cdots, \lfloor (n+1)\alpha + \theta \rfloor - \lfloor n\alpha + \theta \rfloor \Big).$$

Exemple :

$$(1/4, 1/4, 1/4) \in S\Big[(0,0,0), \quad (0,0,1), \quad (0,1,0), \quad (1,0,0)\Big]$$

Démonstration

$p(\alpha, \theta)$ est périodique en θ de période 1 et constante par morceaux prenant au plus $n+1$ valeurs.

$$\int_0^1 \lfloor x + \theta \rfloor d\theta = x \text{ implique } \int_0^1 p(\alpha, \theta) d\theta = (\alpha, \cdots, \alpha).$$

Théorèmes généraux

Théorème

[Altman, Hordijk, G.] Soit $(f_n)_{n\in\mathbb{N}}$ suite de fonctions croissantes. Si

(A) *f_n est multimodulaire pour la base $\mathcal{D}_d$;*

(B) $f_k(a_1, \cdots, a_k) \geqslant f_{k-1}(a_2, \cdots, a_k)$, $\forall k > 1$;

(C) *Pour toute suite $\{a_k\}$ $\exists$ $\{b_k\}$ t.q. $\forall k > m$,*
$f_k(b_1, \cdots, b_{k-m}, a_1, \cdots, a_m) = f_m(a_1, \cdots, a_m)$.

alors le coût à horizon infini

$$g(a) = \overline{\lim}_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} f_n(a_1, \cdots, a_n)$$

est minimisé sur $E_\alpha = \{a \mid \overline{\lim}_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n \geqslant \alpha\}$ en m_α. De plus, $g(m_\alpha)$ est une fonction convexe de α.

Démonstration

Une idée de preuve par analyse convexe ([Jensen]) :

$$
\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}\tilde{f}_n(a_1,\cdots,a_n) &\geqslant \frac{1}{N}\sum_{n=1}^{N}\tilde{f}_k(a_{n-k+1},\cdots,a_n)\text{(par (B))}\\
&\geqslant \tilde{f}_k\left(\frac{1}{N}\sum_{n=1}^{N}a_n,\cdots,\frac{1}{N}\sum_{n=1}^{N}a_n\right)\text{(par convexité)}\\
&\geqslant \tilde{f}(\alpha,\cdots,\alpha)\text{(par croissance)}
\end{aligned}
$$

Propriétés des mots de Sturm (triangulation) :

$$\tilde{f}_k(\alpha,\cdots,\alpha) = f(m_\alpha).$$

La théorie ergodique (théorème ergodique de Weil) permet le passage à la limite.

Théorème général multidimensionnel

Théorème

[Altman G. Hordijk]
Soient $f_n^i(x_1, \ldots, x_n)$ croissantes sur $\mathbb{Z}^n$, pour tout $i \in \{1, \ldots, K\}$ qui vérifient les hypothèses $(A), (B), (C)$. Le coût à horizon infini d'une suite a sur l'alphabet $A = 1 \cdots K$ est

$$g(a) = \overline{\lim_{N \to +\infty}} \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{K} f_n^i(\mathbf{1}_i(a_1, \cdots, a_n))$$

est minimal pour une séquence a qui a des densités $(\alpha_{opt}^1, \cdots \alpha_{opt}^K)$:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_i(a_n) = \alpha_{opt}^i.$$

Cas Particuliers

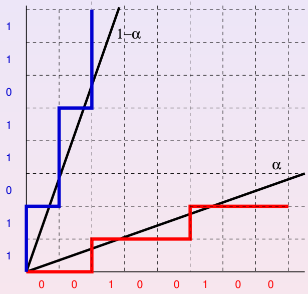
On a bien sûr pour tout $a \in A^{\mathbb{N}}$, $g(a) \geqslant \sum_{i=1}^{K} g^i(m_{\alpha_i})$. Dans quels cas il y a-t-il égalité ?
Dès que $a \in A^{\mathbb{N}}$ est un mot équilibré.
Cas homogène : (f_n^i ne dépend pas de i) la suite qui minimise $g(a)$ est équilibrée, de densité $(1/K, \ldots, 1/K)$. (C'est le "tourniquet".)

Cas $K = 2$: Il existe $\alpha_{opt} \in [0, 1]$ tel qu'une suite optimale est un mot équilibré de densité $(\alpha_{opt}, 1 - \alpha_{opt})$.

Démonstration, cas homogène

Par arguments de symétrie, la densité pour chaque i est la même. Par ailleurs, il est possible de construire une suite équilibrée à K lettres de densités $(1/K, \ldots, 1/K)$.

Démonstration, cas $K = 2$



La figure illustre la propriété suivante : le complémentaire d'un mot équilibré de densité α est un mot équilibré de densité $1 - \alpha$. Cela permet de conclure le cas $K = 2$

Il reste le problème du calcul de α_{opt}.

Exercices

1. Démontrer le cas $K = 2$ en utilisant les théorèmes généraux.
2. Expliquer pourquoi le cas $K \geqslant 3$ pose un problème.

Applications à l' ordonnancement

On considère un protocole TDMA soumis à des perturbations par rafales. Chaque station émitrice est composée de plusieurs réplicas afin d'augmenter la robustesse aux perturbations.
Il reste à placer les réplicas sur un “round” du protocole afin de minimiser la probabilité de perte de tous les paquets transmis par les replicas.
Si les perturbations sont de durées exponentiellement distribuées alors la probabilité de perte est une fonction multimodulaire des placements des paquets. Il faut “répartir” les paquets sur tout le round pour minimiser la probabilité de perte.
Si les perturbations sont de durées distribuées selon une loi de puissance, alors la probabilité de non-perte est multimodulaire. Il faut alors regrouper les réplicas pour minimiser les probabilités de pertes.

Applications à l' ordonnancement (2)

Dnas un contexte temps réel slotté avec des tâches à contraintes strictes et des tâches à contraintes souples. Les tâches à contraintes strictes sont prioritaires sur les tâches souples.
Le temps de réponse moyen des tâches à contraintes souple est une fonction multimodulaire des slots réservés aux tâches souples.
Conséquence : On disperse les exécutions des tâches strictes dans le temps (dans la mesure du possible, c'est à dire en respectant leurs échéances). Cela a pour effet de minimiser le temps de réponse des autres tâches.

Systèmes à événements discrets

Un système à événements discrets (SED) est décrit par :

- un espace d'états et un état X
- une suite d'événements $(z_1, \cdots z_n, \cdots)$ à valeur dans $\mathcal{Z}$ (souvent inclus dans $\mathbb{R}^k$)
- une *dynamique* D donnant la valeur de l'état en réponse à la suite d'événements :
  évolution jusqu'à l'étape n : $X_n = D_n(z_1, \cdots z_n)$.

SED *stochastique* si les événements $(z_1, \cdots z_n, \cdots)$ forment un processus aléatoire (les innovations).

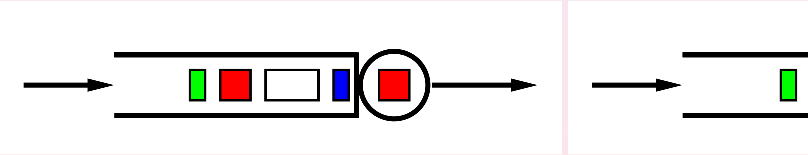
Exemples :

- files d'attente
- réseaux de Petri
- automates
- ...

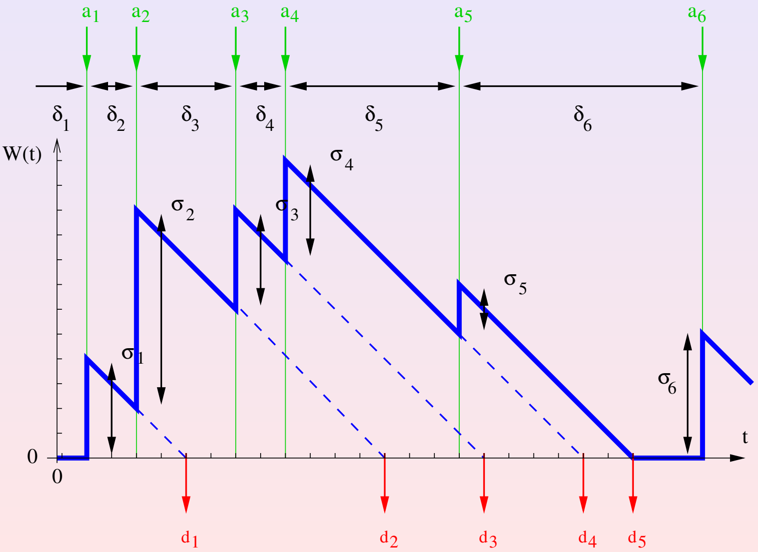
Files d'attente

La file d'attente générique : G/G/1//FIFO (ou : G/G/1)

- Processus d'arrivée quelconque
- Distribution de la durée des services générale
- 1 serveur, service FIFO
- capacité de stockage de clients infinie.



Files d'attente : courbe de charge



Équation de Lindley

Représentation de l'évolution de la file d'attente FIFO.
Choix de l'état : $D_n = W_n$ le temps d'attente du n-ème client.
Le temps d'attente égale la quantité de travail en instance à la date d'arrivée.
L'innovation est : $z_n = (\sigma_n, \delta_n) \in (\mathbb{R}_+)^2$

$$W_n \;=\; \max\left(W_{n-1} + \sigma_{n-1} - \delta_n, 0\right) \;.$$

Équation de Lindley.

SED contrôlé

Soit $(a_1, \cdots, a_n, ...)$, la suite des actions du contrôleur à valeur dans $\mathcal{A}$. La dynamique du système est modifiée par le contrôle :

$$D_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n)$$

Exemple : contrôle d'admission dans la file G/G/1, actions dans $\mathcal{A} = \{0, 1\}$.
$a_i = 1$: le client i est accepté et entre dans la file.
$a_i = 0$: le client i est rejeté définitivement.
La charge juste avant l'arrivée du n-ème client est :

$$W_n \;=\; \max\left(W_{n-1} + a_{n-1}\sigma_{n-1} - \delta_n, 0\right)\ .$$

Objectif du contrôle

Le contrôleur essaie de minimiser une fonction de coût G qui est construite à partir de coûts immédiats : $g_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n)$.

Coût moyen à horizon fini (N) : $G = \frac{1}{N} \mathbb{E} \sum_{n=1}^{N} g_n.$

Coût moyen à horizon infini : $G = \overline{\lim_{N \to \infty}} \frac{1}{N} \mathbb{E} \sum_{n=1}^{N} g_n.$

Coût actualisé : $G = \overline{\lim_{N \to \infty}} \mathbb{E} \sum_{n=1}^{N} \alpha^n g_n.$

Structure d'Information

Quelle information est disponible pour le contrôleur ?
Fonction d'information : $Y_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n)$.
En boucle fermée : Information totale
$Y_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n) = (a_1, \cdots, a_{n-1}, z_1, \cdots, z_n)$.
En feedback d'état : Information sur l'état courant
$Y_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n) = D_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n)$.
Information en boucle ouverte :
$Y_n(a_1, \cdots, a_{n-1}, z_1, \cdots, z_n) = (a_1, \cdots, a_{n-1})$.

Dans un contexte probabiliste, le concept d'information est souvent plutôt lié à la mesurabilité de la politique optimale.

Politiques

La **politique** de contrôle (ou stratégie) est la règle de décision qui produit une action à partir de l'information disponible.

Politique : $u = (u_1, \cdots, u_n, \cdots)$ avec $u_n : Y_n \mapsto \mathcal{A}$.

Une politique optimale u^* est telle que $G(u^*)$ est minimale parmi les $G(u)$ obtenus pour les politiques qui ont la même information à leur disposition.

Une politique est donc optimale pour un critère donné et une structure d'information donnée.

Contrôle d'admission dans un réseau

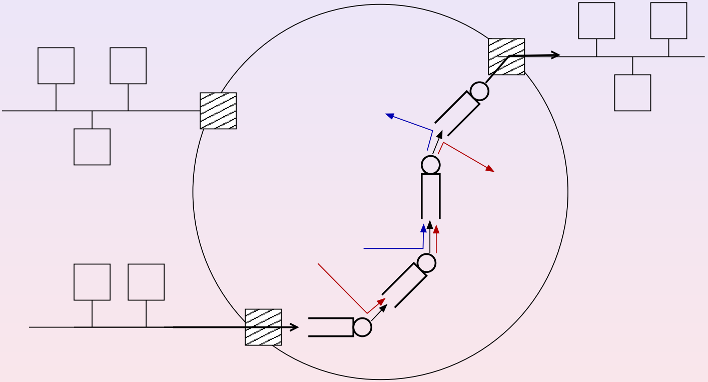
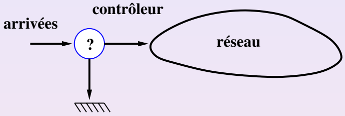


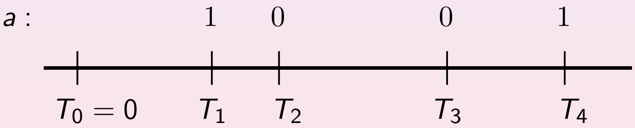
Fig.: Une connexion et le trafic transverse

Suite de contrôle

Problème du contrôleur :



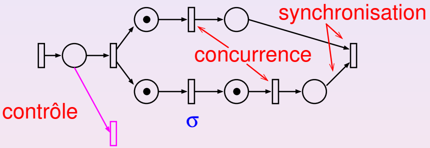
Le contrôle est représenté par une suite binaire infinie $\{a_i, \quad i \in \mathbb{N}\}$.
$a_i = 1$: le paquet i est accepté et entre dans le réseau.
$a_i = 0$: le paquet i est rejeté définitivement.



But : trouver la séquence d'admission en boucle ouverte qui optimise les délais dans le réseau (fonction de coût) sans rejeter trop de paquets (contrainte).

Modèle du réseau

Le réseau est modélisé par un réseau de Petri (graphe d'événements) avec temporisations stochastiques, $R = (P, Q, q_0, G, M_0, \Phi, \tau)$.



P : ensemble de places
Q : ensemble des transitions
q_0 : transition d'entrée
G : connections places-transitions
M_0 : marquage initial
$\Phi = \{\phi_q(n)\}_{q \in Q, n \in \mathbb{N}}$: temporisation de transitions
$\tau = \{\tau_n\}_{n \in \mathbb{N}}$: dates d' arrivées.

Hypothèses stochastiques : τ_n est un processus ponctuel stationnaire, les $\phi_q(n)$ sont stationnaires pour tout q. Les τ_n sont indépendants des $\phi_q(n)$.

Théorèmes généraux (2)

Théorème

[Altman, Hordijk, G.]

Soit $\mathcal{S} = (P, Q, q_0, G, M_0, \Phi, \tau)$ un graphe d'événements. Soit $W_n^i(a_1, \cdots, a_n)$ le temps de trajet jusqu'au nœud i au moment de la n-ième arrivée. La fonction $\mathbb{E}h \circ W_n^i(a_1, \cdots, a_n)$ est multimodulaire sur $\mathcal{D}_d$ si h est convexe croissante.

Théorème

[Altman, Hordijk, G.]

i- $\mathbb{E}W_n^i(a_1, \cdots, a_n)$ est croissante

ii- $\mathbb{E}W_n^i(a_1, \cdots, a_n) = \mathbb{E}W_m^i(0, \cdots, 0, a_1, \cdots, a_n)$, $n < m$.

iii- $\mathbb{E}W_n^i(a_{m-n+1}, \cdots, a_m) \leqslant \mathbb{E}W_m^i(a_1, \cdots, a_m)$, $n < m$.

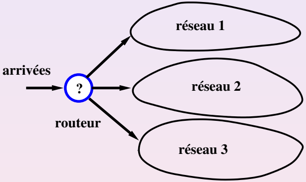
Conclusion : la séquence d'admission en boucle ouverte recherchée est équilibrée.

Exercices

1. Au vu de la figure sur le contrôle d'admission avec trafic transverse, expliquer en quoi les hypothèses stochastiques sont restrictives.
2. Démontrer le théorème 3.

Problèmes de routages

Le contrôle est une suite $\{a_n\}_{n\in\mathbb{N}}$ à valeurs dans $\{1, \cdots, K\}$.
Si $a_n = i$ le paquet n est dirigé vers le réseau S_i.

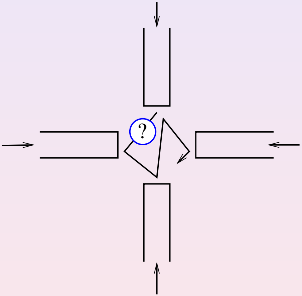


Sous les hypothèses très générales sur les caractéristiques de ces réseaux vues dans le cas de l'admission, le routage optimal est équilibré dans le cas $K = 2$ (et très difficile à caractériser sinon). (On applique le théorème général multidimensionnel avec $K = 2$.)
Il reste le problème du calcul de la densité du routage optimal.

Problèmes de Polling

Polling : un serveur doit choisir, entre plusieurs files, laquelle servir.

C'est le problème dual de celui du routage.



À nouveau le théorème multidimensionnel s'applique (avec quelques précautions). Avec $K = 2$, la politique de polling optimale est équilibrée. Là encore le problème du calcul de la densité optimale se pose.

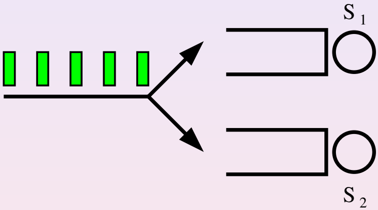
Problèmes de calcul

À ce stade, il est su que la politique optimale est équilibrée.
Mais la densité optimale du routage/de l'admission n'est pas connue. Elle dépend *a priori* des propriétés statistiques du processus d'arrivées et des services dans le réseau.

Calcul de cette densité optimale pour le routage entre deux files, dans deux cas :

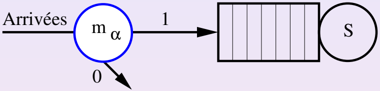
- arrivées périodiques et services constants
  $\longrightarrow$ calcul exact *via* factorisation de mots mécaniques
- arrivées selon un processus de Poisson et services de distribution exponentielle.
  $\longrightarrow$ calcul numérique *via* chaînes de Markov.

Le modèle déterministe

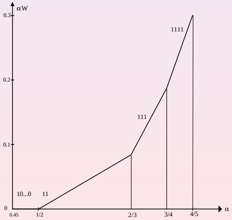


Hypothèses : une arrivée de paquet toutes les secondes. Les temps de service S_1 et S_2 dans les deux serveurs sont constants.

Temps d'attente moyen dans 1 file



Temps d'attente moyen pondéré dans une file : $\alpha W(m_\alpha)$ pour $S = 5/4$ ($1/S = \langle 2, 2, 2, 2\rangle$).



Convexité du temps d'attente

Soit α_1 et α_2 deux nombres rationnels dans l'intervalle de stabilité : $J = [1 - 1/S_2, 1/S_1]$.
Soit $m_1 = m_{\alpha_1}$, $m_2 = m_{\alpha_2}$ et $m = m_1^{q_2} m_2^{q_1}$.
Soit $N(m)$ le nombre moyen de clients dans la file durant m : on a $N(m) = \frac{q_2 N(m_1) + q_1 N(m_2)}{q_1 + q_2}$.
Soit $\alpha = \frac{q_2 \alpha_1 + q_1 \alpha_2}{q_1 + q_2}$,
par le théorème général d'optimisation,
$N(m_\alpha) \leqslant N(m) = \frac{q_2 N(m_1) + q_1 N(m_2)}{q_1 + q_2}$.
Et donc $\forall \lambda \in [0, 1]$

$$N(m_{\lambda\alpha_1 + (1-\lambda)\alpha_2}) \leqslant \lambda N(m_{\alpha_1}) + (1 - \lambda) N(m_{\alpha_2}).$$

Par la formule de Little $N = \alpha(W + S)$.

Linéarité sur les intervalles de Farey

Un intervalle de Farey est un intervalle $I =]a, b[$ tel que tout nombre rationnel dans I a un dénominateur plus grand que ceux de a et b.
Par exemple $I =]1/2, 1[$ et $I =]1/3, 2/5[$ sont des intervallles de Farey
Si on regarde les mots équilibrés associés : $m_{1/3} = 100$ et $m_{2/5} = 10100$, on remarque que $m_{2/5}m_{1/3} = 10100100$ est un mot équilibré (de densité (1+2)/(3+5)). En fait cette propriété est vraie pour tout intervalle de Farey.
De proche en proche, le mot équilibré m_α de densité $\alpha \in I$ peut se factoriser en m_{a_1} et m_{a_2}.
Cela montre la linéarité de $N(m_\alpha)$ sur l'intervalle I si I est inclus dans l'intervalle de stabilité J et aussi de $\alpha W(m_\alpha)$ par la formule de Little.

Convergents de la fraction continue.

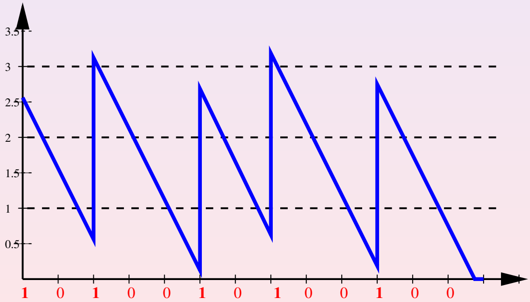
Pour une file, si on considère les convergents de la fraction continue par en-dessous de son intensité de service.
Deux convergents successifs forment un intervalle de Farey sur lequel le temps moyen d'attente est linéaire.
Pour des arrivées formées de mots équilibrés de densités égales aux convergents successifs, la charge du système ne s'annule jamais avant la fin. Cela permet de donner une formule close du temps d'attente en ces points et donc de calculer complètement la fonction de coût :

$$\mathbb{E}W(a) = \Pr(\text{routé vers } S_1)\, \mathbb{E}W_1(a) + \Pr(\text{routé vers } S_2)\, \mathbb{E}W_2(1-a).$$

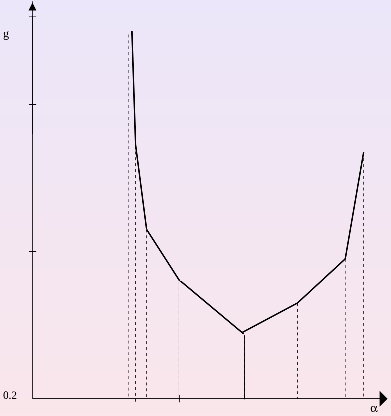
On en déduit que $\alpha W(m_\alpha) + (1-\alpha)W(m_{1-\alpha})$ est minimal en un convergent de $1/S_1$ ou de $1/S_2$.

Charge sous un convergent

On considère une file d'attente avec un temps de service $S = 51/20$. La suite des convergents par en-dessous de $1/S$ est : $1/3, 3/8, 5/13, \cdots, 20/51$. Si on choisit $m_{5/13} = 1010010100100$ on obtient la forme suivante de la charge dans la file.



Courbe du temps de réponse moyen total

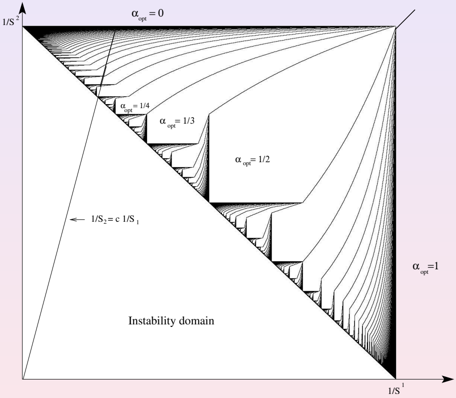


Recherche de $\alpha_{opt} = \arg\min_{\alpha \in I_s} \alpha W_1(m_\alpha) + (1-\alpha) W_2(m_{1-\alpha})$.

L'optimum α_{opt} est rationnel ! ! !

La politique optimale est donc périodique !

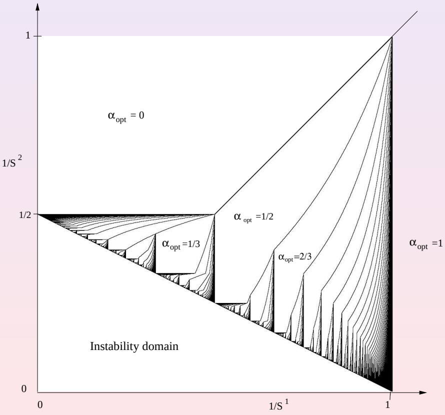
Taux optimal dans deux files



On cherche
$\alpha_{opt} = \arg\min_{\alpha \in I_s} \alpha(W_1(m_\alpha) + S_1) + (1-\alpha)(W_2(m_{1-\alpha}) + S_2)$.

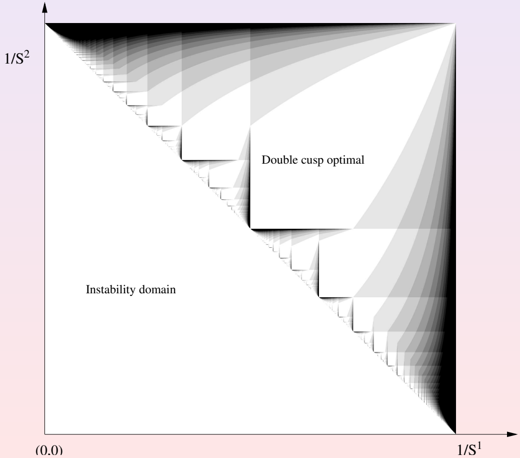
Taux optimal dans $N_1 + N_2$ files

$$\alpha_{opt} = \arg\min_{\alpha \in I_s} = N_1 \alpha R^1(m_\alpha) + \frac{(1-\alpha N_1)}{N_2} R^2(m_{(1-\alpha N_1)/N_2}),$$



Comportement du taux optimal

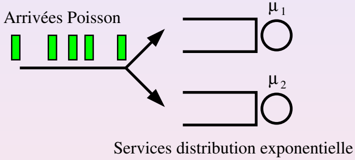
$\alpha_{opt} = \arg\min_{\alpha \in I_s} = N_1 \alpha R^1(m_\alpha) + \frac{(1-\alpha N_1)}{N_2} R^2(m_{(1-\alpha N_1)/N_2})$.
Le niveau de gris donne l'ordre de grandeur du dénominateur. Plus de 95% des points ont un dénominateur inférieur à 100.



Exercices

- (*) Montrer la propriété de factorisation des mots équilibrés de densités dans un intervalle de Farey
- (*) Montrer la formule de Little dans le cas présent.

Problème de routage, cas stochastique



On considère le problème de routage dans deux files avec des temps de services aléatoires, exponentiellement distribués de paramètres μ_1 et μ_2. Les arrivées forment un processus de Poisson d'intensité λ.

Quasi-processus de naissance et de mort

Le comportement d'une file avec une politique de routage ℓ-périodique peut être vu comme une chaîne de Markov en temps continu sur $\mathbb{N} \times \{1, \cdots \ell\}$.

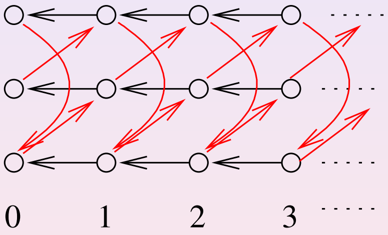


FIG.: Le générateur Q de la chaîne, avec la suite de routage $m_{2/3} = 110$.

Le calcul de la probabilité stationnaire de cette chaîne peut être fait numériquement par la technique du noyau (série génératrice + analyse complexe) ou la technique des matrices géométriques [14].

Suites de routage optimales

Comme dans le cas déterministe, le routage optimal est une suite équilibrée de densité α_{opt} qu'il s'agit de calculer.

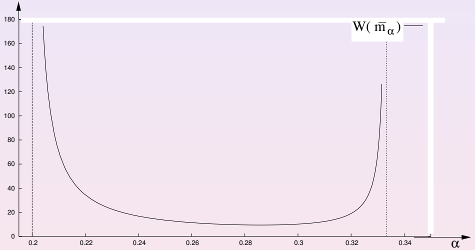


FIG.: La fonction $\alpha \mapsto W(m_\alpha)$ est convexe en α [Hordijk, van der Laan 2003].

On peut utiliser les techniques numériques classiques de minimisation de fonctions convexes pour trouver α_{opt}.

Étude de α_{opt}

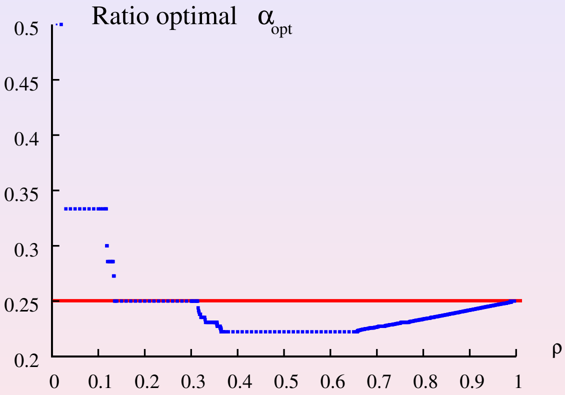
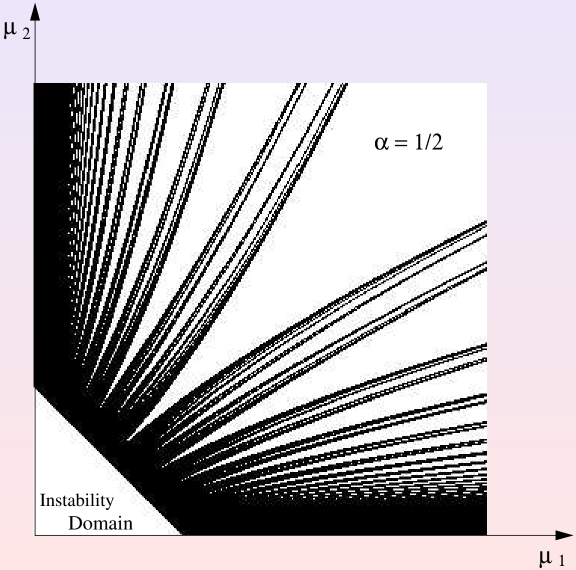


FIG.: La fonction $\rho \mapsto \alpha_{opt}$ avec $\mu_1 = 7/16$ et $\mu_2 = 21/16$

En faisant varier λ afin que la charge globale ρ varie de 0 à 1, le comportement de α_{opt} est très irrégulier (continu ?).

Étude de α_{opt} (II)



Comparaison avec des routages classiques

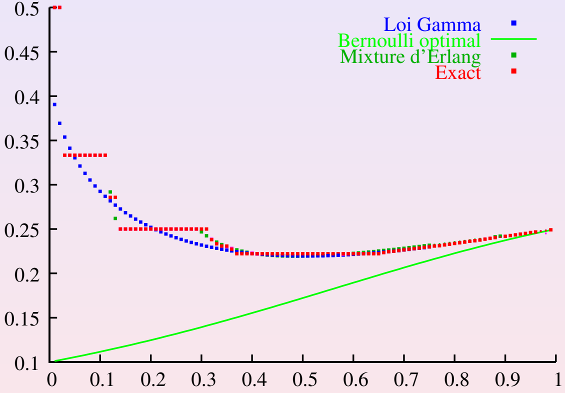


FIG.: Comparaisons des taux optimaux des routages Bernoulli, Gamma, mixture d'Erlang et équilibrés

Cas du polling

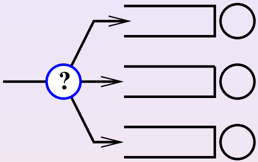
L'allocation optimale dans les systèmes de polling à deux files (dans le cas déterministe ou stochastique) montre le même comportement que celui du routage optimal.

Extensions

Il existe des extensions de ces résultats dans plusieurs directions.

- Ordres partiels et déséquilibre dans les cas multidimensionnels.
- Contrôle optimal en boucle fermée.
- Multimodularité en espace [9].

Routage mutidimensionel

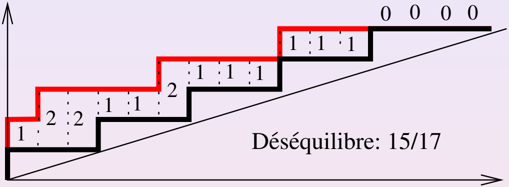


La fonction de coût est la somme des charges moyennes de chaque file. $W(a) = \sum_{i=1}^{K} W_i(\mathbf{1}_i(a))$ Alors pour tout a de densité $(\alpha_1, \cdots, \alpha_K)$,

$$0 \leqslant W(a) - \sum_{i=1}^{K} W_i(m_{\alpha_i}) \leqslant \overline{\delta} U(a),$$

où $U(a)$ est le déséquilibre de a (la somme des déséquilibres pour chaque $i \in \{1 \cdots, K\}$) et $\overline{\delta}$ est la moyenne des interarrivées.

Déséquilibre



Le déséquilibre d'une suite binaire de densité α se mesure comme un "écart" au mot de Strum de même densité. Le déséquilibre d'une suite multidimensionnelle est la somme des déséquilibres de toutes ses lettres.

De plus, en toute dimension K, il existe des suites billards mutidimensionnelles de déséquilibre $K/2 - 1$ [17].

Contrôle optimal en boucle fermée

Théorème

Si la fonction de coût est multimodulaire, alors la politique optimale est monotone.

Application au routage dans plusieurs files d' attentes homogènes : avec information totale et des hypothèses supplémentaires sur les processus de service et d' inter-arrivées (IFR),la politique "join the shortest queue" est optimale.

Notes bibliographiques

Les mots de Sturm et les mots crochets ont été introduits par Morse et Hedlund. Ils sont utilisés dans plusieurs domaines scientifique (physique statistique, image de synthèse, combinatoire, systèmes dynamiques, analyse convexe et en optimisation). Une présentation très bien faite de leurs propriétés combinatoires est faite dans [13]. La conjecture de Fraenkel est étudiée dans [16, 1].

La multimodularité a été introduite par Bruce Hajek [12] puis généralisée par la suite dans [2] et dans [15]. La multimodularité spatiale est utilisée dans [10] pour résoudre des problèmes de contrôle optimal.

La représentation des réseaux de communication par des équations d'évolution matricielles dans l'algèbre (max,plus) est développée dans [3], ainsi que dans un grand nombres d'articles par la suite. À cet égard, on peut aussi citer les travaux autour du "Network Calculus", qui généralise cette approche aux équations fonctionnelles dans (min,plus) [5, 4].

Enfin, l'optimisation en boucle ouverte du routage et du polling est développée dans une série d'articles [7, 6, 8].

[1] E. Altman, B. Gaujal, and A. Hordijk.
Balanced sequences and optimal routing.
Journal of the ACM, 47(4) :752–775, 2000.

[2] E. Altman, B. Gaujal, and A. Hordijk.
Discrete-Event Control of Stochastic Networks : Multimodularity and Regularity.
Number 1829 in LNM. Springer-Verlag, 2003.

[3] F. Baccelli, G. Gohen, G. J. Olsder, and J.-P. Quadrat.
Synchronization and Linearity.
Wiley, 1992.

[4] J.Y. Le Boudec and P. Thiran.
Network calculus, A theory of deterministic queueing systems for the internet.
Number 2050 in LNCS. Springer Verlag, 2000.

[5] C.S. Chang.
Performance Garanties in communication networks.

Springer, 2000.

[6] B. Gaujal, A. Horidjk, and D. Van der Laan.
On the optimal policy for deterministic and exponential polling systems.
Technical report, INRIA, 2005.

[7] B. Gaujal and E. Hyon.
Optimal routing policy in two deterministic queues.
Réseaux et Systèmes Répartis – Calculateurs Parallèles, 13, 2001.
Also available as INRIA RR-3997.

[8] B. Gaujal, E. Hyon, and A. Jean-Marie.
Optimal routing in two parallel queues with exponential service times.
In *WODES*, pages 193–198, Reims, France, 2004. IFAC.
A long version is available in
http ://www.inria.fr/rrrt/rr-5109.html.

[9] P. Glasserman and D. D. Yao.
Monotone optimal control of permutable GSMPs.
Mathematics of Operation Research, 19 :449–476, 1994.

[10] P. Glasserman and D. D. Yao.
Monotone Structure in Discrete Event Systems.
Wiley, 1994.

[11] R. Graham.
Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta] : n = 1, 2, \cdots\}$.
Journal of Combinatorial Theory, 15 :354–358, 1973.

[12] B. Hajek.
Extremal splittings of point processes.
Mathematics of Operation Research, 10(4) :543–556, 1985.

[13] M. Lothaire.
Mots, chapter Tracé de droites, fractions continues et morphismes itérés (J. Berstel).

Hermes, 1991.

[14] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts.
A single-server queue with server vacations and a class of non-renewal arrival processes.
Advances in Applied Probability, 22 :275–705, 1990.

[15] K. Murota.
Discrete convex analysis.
Monographs on Discrete Mathematics and Applications. SIAM, 2003.

[16] R. Tijdeman.
On complementary triples of Sturmian sequences.
Indagationes Mathematicae, pages 419–424, 1996.

[17] D. A. van der Laan.
Routing jobs to servers with deterministic service times.
Technical Report MI N. 2000-20, Leiden University, 2000.