

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de la Technologie
Département de Mathématiques et Informatique



Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en : Mathématiques
Domaine : Mathématiques et Informatique
Filière : Mathématiques
Spécialité : Probabilité et Statistiques Appliquées
Thème

Analyse d'un réseau fluide multi-classe

Présenté Par :

1) Melle Medraoui Manel

Devant le jury composé de :

Mme GAOUAR Soumia	MCB	UAT.B.B (Ain Temouchent)	Présidente
Mme MESSABIHI Aicha	MCB	UAT.B.B (Ain Temouchent)	Examinatrice
Mme SAKHI Hanane	MCB	UAT.B.B (Ain Temouchent)	Encadrante

Année Universitaire 2023/2024

Remerciements

J'aimerais en premier lieu remercier ALLAH qui m'a donné la volonté et le courage pour achever ce travail.

Je tiens à remercier tout d'abord ma directrice de mémoire Dr. SAKHI HANANE qui m'a proposé le sujet.

Je la remercie de m'avoir encadré, orienté, aidé et conseillé .

Je tiens à remercier aussi les membres de jury d'avoir accepté l'évaluation de ce travail.

J'adresse un grand merci à toute ma famille qui a toujours été présente lorsque j'en ai eu besoin, en particulier mon père, ma mère ,mes surs et frères pour leurs encouragements.

Enfin, je remercie toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

Dédicace

Je dédie cet humble et modeste travail avec grand amour, sincérité et fierté :

A mes chers parents, source de tendresse, et de noblesse. Puisse cette étape constituer pour vous un motif de satisfaction.

A mes frère et mes soeurs, en témoignage de la fraternité, avec souhaits de bonheur, de santé et de succès.

Et à tous les membres de Ma famille.

A tous mes amie, tous mes professeurs

Table des matières

Introduction	2
1 Notations et définitions générales	4
1.1 Rappels sur les chaînes de markov	4
1.2 Files d'attente	7
1.3 Réseau de files d'attente	8
1.3.1 Différentes disciplines de service	8
1.3.2 Processus de Markov sous-jacent	10
1.3.3 Equations des réseaux de files d'attente	11
2 Stabilité des modèles stochastiques	15
2.1 Heuristique	15
2.2 Résultat généraux	16
2.2.1 Ergodicité d'une chaîne de Markov	17
2.2.2 Transience d'une chaîne de Markov	20
2.2.3 Propriétés sur les temps de retour d'une chaîne de Markov	20
2.3 Limites fluides	25
2.3.1 Renormalisation des processus	26
2.3.2 Limites fluides	26
2.3.3 Relation entre renormalisation et ergodicité	27
2.3.4 Approximation fluide et stabilité	28
3 Conditions de stabilité d'un réseau de file d'attente multiclasse avec priorité	30
3.1 Description du modèle et notations	31
3.2 Résultat principal	34
Conclusion	39
Annexe	40
Bibliographie	42

Introduction

L'organisation des "réseaux de télécommunications" vise à gérer les phénomènes qui se produisent lors du traitement des communications qui y sont traitées et plus généralement, les "services" qui y sont fournis. Ces phénomènes sont régis par l'apparition aléatoire des demandes et sont étudiés indépendamment du choix de la technologie utilisée. Ceux-ci suivent le formalisme de la théorie des probabilités, qui a donné naissance au concept de "intensité de trafic" et jouera un rôle central dans sa compréhension.

Le concept de réseaux de files d'attente a été initialement présenté par Scherr dans les années 1960, lorsqu'il a examiné les retards et le temps d'attente dans les réseaux informatiques. Au cours des années 1970, les réseaux de files d'attente ont été utilisés par les chercheurs pour modéliser les systèmes de fabrication flexibles. L'utilisation de la théorie des files d'attente était très limitée au début, mais au fil des années, ce domaine de la recherche a beaucoup évolué et il a inspiré de nombreux chercheurs.

Le domaine de recherche est la stabilité du trafic et du stockage dans les réseaux informatiques. Le réseau est constitué d'un ou plusieurs serveurs. Les flux de messages provenant de différentes catégories arrivent et sont traités par les serveurs individuellement. Ces messages sont ensuite placés dans différentes files d'attente du réseau.

Les modèles de réseaux de file d'attente ont été appliqués avec succès pour évaluer et optimiser divers systèmes, notamment les systèmes informatiques, les réseaux de communication, les serveurs Web, les systèmes de fabrication ainsi que les trafics routiers peuvent être représentés au moyen de modèles de files d'attente.

On dit qu'un réseau est stable lorsque la congestion initiale est stable. Il y'a deux façons de connaître le comportement du réseau. L'une consiste à étudier le processus $(X(t))$, (où $X(t)$ est le processus de Markov de sauts représentant le modèle de réseau de file d'attente) en régime permanent, et l'autre à étudier le comportement asymptotique de ce réseau lorsque certains paramètres sont augmentés (limites de fluide) .

En général, l'étude du processus $(X(t))$ en régime permanent est difficile. Dans la plupart des modèles décrivant le comportement à l'équilibre, la loi du régime permanent est exprimée sous la forme du produit bien connu. La méthode de la limite fluide est utilisée pour étudier le comportement asymptotique du réseau. La méthode modifie le processus en accélérant le temps et en modifiant l'espace d'un certain paramètre, et étudie le comportement du processus modifié lorsque

ce paramètre tend vers l'infini.

L'objectif de notre travail est d'analyser la stabilité d'un réseau de file d'attente fluide multi-classes. Ce mémoire se compose de trois chapitres organisés de la manière suivante :

Dans le premier chapitre, on introduit des notations et définitions générales sur les chaînes de Markov, les files d'attente et réseau de files d'attente et ses disciplines. Le deuxième chapitre est consacré à stabilité des modèles stochastiques, on utilise les limites fluides et la fonction Lyapunov comme principale outil dans le domaine. Ensuite, les principaux critères d'ergodicité et transience d'une chaîne de Markov.

Dans le troisième chapitre, on étudie les conditions de stabilité d'un réseau de file d'attente fluide multi-classe avec priorité.

En finalisant notre mémoire par une conclusion et quelques références bibliographiques.

Notations et définitions générales

Les chaînes de Markov et les files d'attente sont deux concepts importants en théorie des probabilités et en modélisation des systèmes. Bien qu'ils soient distincts, ils sont souvent utilisés ensemble pour modéliser et analyser les systèmes de files d'attente.

1.1 Rappels sur les chaînes de markov

Nous considérons dans ce chapitre une suite de variables aléatoires $X = \{X_n, n \in \mathbb{N}\}$ définies sur un espace de probabilité (Ω, \mathcal{F}, P) , avec des valeurs dans un ensemble dénombrable S et satisfaisant la propriété de Markov, c'est-à-dire que le passé et le futur de X sont indépendants lorsque son état actuel est connu, X_n ne dépend que de X_{n-1} . Le temps est représenté ici par l'indice n , ce qui explique que l'on parle de temps discret. L'ensemble S est souvent appelé l'espace d'état (voir [22]).

Définition 1.1. *Un processus stochastique $X = \{X_n, n \in \mathbb{N}\}$ sur un espace d'états S est une chaîne de Markov à temps discret si :*

1. $\forall n \geq 0, X_n \in S,$
2. $\forall n \geq 1$ et $\forall i_0, \dots, i_{n-1}, i_n \in S,$ nous avons :

$$P \{X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} = P \{X_n = i_n \mid X_{n-1} = i_{n-1}\}.$$

Définition 1.2. *Une chaîne de Markov à temps discret $X = X_n, n \in \mathbb{N}$ sur un espace d'états dénombrable S est dite homogène si, $\forall n, k \in \mathbb{N}$ et $\forall i, j \in S,$ nous avons :*

$$P \{X_{n+k} = j \mid X_k = i\} = P \{X_n = j \mid X_0 = i\}.$$

Définition 1.3. *On considère, $\forall i, j \in S, P_{i,j} = P \{X_n = j \mid X_{n-1} = i\}$ et on définit la matrice de probabilité de transition P de la chaîne de Markov X_n comme suit :*

$$P = (P_{i,j})_{i,j \in S}$$

Propriétés 1.1.1. *On a alors, par définition, $\forall i, j \in S,$*

$$P_{i,j} \geq 0 \text{ et } \sum_{j \in S} P_{i,j} = 1.$$

Une matrice pour laquelle ces deux propriétés se vérifient est appelée matrice stochastique. Pour tout $n \in \mathbb{N}$, nous écrivons $(P^n)_{i,j}$ le coefficient (i, j) de la matrice P^n , où nous définissons $P^0 = I$, avec I la matrice identité dont la dimension sera vérifiées, donnée plus loin ici, elle est égale au nombre d'états $|S|$ de S . Nous écrivons $\alpha = (\alpha_i, i \in S)$ le vecteur de lignes contenant la distribution initiale de la chaîne de Markov X_n , définie par :

$$\alpha_i = P \{X_0 = i\}.$$

$\forall i \in S$, on a donc :

$$\alpha_i \geq 0 \text{ et } \sum_{i \in S} \alpha_i = 1.$$

Définition 1.4. Une variable aléatoire T à valeurs dans $\mathbb{N} \cup \{\infty\}$ est appelée temps d'arrêt du processus $X, \forall n \geq 0, \{T = n\} \in \mathcal{F}_n$.

Remarque 1. nous utilisons souvent la variable $\tau(j)$ qui compte le nombre de transitions nécessaires pour atteindre l'état j , définie par :

$$\tau(j) = \inf \{n \geq 1 \mid X_n = j\},$$

où $\tau(j) = \infty$ si cet ensemble est vide. $\forall j \in S, \tau(j)$ est un temps d'arrêt puisque

$$\{\tau(j) = 0\} = \emptyset \in \mathcal{F}_0, \{\tau(j) = 1\} = \{X_1 = j\} \in \mathcal{F}_1$$

et, $\forall n \geq 2$,

$$\{\tau(j) = n\} = \{X_n = j, X_k \neq j, 1 \leq k \leq n-1\} \in \mathcal{F}_n.$$

Soit T un temps d'arrêt, et \mathcal{F}_T la σ -algèbre d'événements exprimée comme une fonction de X_0, \dots, X_T , c'est-à-dire :

$$\mathcal{F}_T = \{B \in \mathcal{F} \mid \forall n \in \mathbb{N}, B \cap \{T = n\} \in \mathcal{F}_n\}.$$

$\forall i, j \in S$, nous définissons $f_{i,j}$ comme

$$f_{i,j} = P \{\tau(j) < \infty \mid X_0 = i\} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}.$$

La quantité $f_{i,i}$ est la probabilité, à partir de i , que le premier retour à l'état i se produise en un temps fini et, pour $i \neq j, f_{i,j}$ est la probabilité, à partir de i , que la première visite à l'état j se produise en un temps fini.

Définition 1.5. Un état $i \in S$ est dit récurrent si $f_{i,i} = 1$ et transitoire si $f_{i,i} < 1$. Une chaîne de Markov est dite récurrente (respectivement transitoire) si tous ses états sont récurrents (respectivement transitoires).

Définition 1.6. Un état $i \in S$ est dit absorbant si $P_{i,i} = 1$.

Tous les états absorbants sont récurrents. En effet, si i est un état absorbant alors, par définition, nous avons $f_{i,i}^{(n)} = \mathbb{1}_{\{n=1\}}$ et donc $f_{i,i} = 1$, ce qui signifie que l'état i est récurrent.

Définition 1.7. Un état $j \in S$ est dit accessible à partir d'un état $i \in S$ s'il existe un entier $n \geq 0$, tel que $(P^n)_{i,j} > 0$. On écrit alors $i \rightarrow j$.

Définition 1.8. On dit que deux états i et j communiquent s'ils sont accessibles l'un à l'autre. On écrit alors $i \longleftrightarrow j$.

Théorème 1.1.1. Cette relation de communication est une relation d'équivalence, ce qui signifie que $\forall i, j, k \in S$, on a :

1. $i \longleftrightarrow i$ (réflexivité).
2. $i \longleftrightarrow j \iff j \longleftrightarrow i$ (symétrique).
3. $i \longleftrightarrow j, j \longleftrightarrow k \implies i \longleftrightarrow k$ (transitivité).

Démonstration. Chaque état i est accessible à partir de lui-même car $(P^0)_{i,i} = 1 > 0$. La relation est donc réflexive. Elle est également symétrique, par définition. Quant à la transitivité, si $i \longleftrightarrow j$ et $j \longleftrightarrow k$ alors il existe des entiers n et m tels que $(P^n)_{i,j} > 0$ et $(P^m)_{j,k} > 0$. On obtient alors :

$$(P^{n+m})_{i,k} = \sum_{l \in S} (P^n)_{i,l} (P^m)_{l,k} \geq (P^n)_{i,j} (P^m)_{j,k} > 0$$

On a donc prouvé que $i \longrightarrow k$. De la même façon, on prouve que $k \longrightarrow i$. \square

Comme pour toute relation d'équivalence, les classes d'équivalence forment une partition de l'espace d'état S , c'est-à-dire que leurs classes sont des classes d'équivalence. l'espace d'états S , c'est-à-dire que leurs classes ne sont pas vides et disjointes et que leur union est égale à S . Une classe d'équivalence regroupe tous les états communicants. $\forall i \in S$, la classe d'équivalence $C(i)$ de l'état i est définie par :

$$C(i) = \{j \in S \mid i \longleftrightarrow j\}.$$

Ces classes d'équivalence sont les composantes fortement connectées du graphe liées à la chaîne de Markov X .

Définition 1.9. Une chaîne de Markov X est dite irréductible si tous ses états communiquent, c'est-à-dire si elle possède une seule classe d'équivalence ou si son graphe admet une seule composante fortement connectée.

$$i \longleftrightarrow j, \forall i, j \in S$$

Propriétés 1.1.2. $\forall i, j \in S$, on a :

1. $i \longleftrightarrow j$ et j récurrent $\implies i$ récurrent et $f_{i,j} = f_{j,i} = 1$.
2. $i \longleftrightarrow j$ et j transient $\implies i$ transient.

Définition 1.10. Un sous-ensemble non vide C d'états de S est dit fermé (Close) si $\forall i \in C$ et, $\forall j \notin C$, on a $P_{i,j} = 0$. Rappelons que pour tout sous-ensemble C d'états de S , on a :

$$\sum_{j \in C} (P^n)_{i,j} = P \{X_n \in C \mid X_0 = i\}.$$

Théorème 1.1.2. Si X est une chaîne de Markov irréductible et récurrente, alors $\forall i, j \in S$, on a $f_{i,j} = 1$ et $P \{\tau(j) < \infty\} = 1$.

Démonstration. Si la chaîne X est irréductible et récurrente, on a, d'après (1.1.2), $f_{i,j} = P \{\tau(j) < \infty \mid X_0 = i\} = 1$ pour tout $i, j \in S$.

On en déduit alors que, $\forall j \in S$,

$$P \{\tau(j) < \infty\} = \sum_{i \in S} P \{X_0 = i\} P \{\tau(j) < \infty \mid X_0 = i\} = \sum_{i \in S} P \{X_0 = i\} = 1$$

ce qui complète la démonstration. \square

Définition 1.11. Nous appelons mesure sur S tout vecteur de ligne $\pi = (\pi_j, j \in S)$ tel que que $0 \leq \pi_j < 1$. Nous disons que la chaîne de Markov X a une mesure invariante π si π est une mesure sur S et si $\pi = \pi P$. La mesure π est dite positive si $\pi_j > 0, \forall j \in S$.

Définition 1.12. La période $d(i)$ d'un état $i \in S$ est définie par :

$$d(i) = \text{pgcd} \{n \geq 1 \mid (P^n)_{i,i} > 0\},$$

en utilisant la convention $d(i) = 0$ si $(P^n)_{i,i} = 0, \forall n \geq 1$. Si $d(i) = 1$, l'état i est dit apériodique.

Les chaînes de Markov offrent un cadre mathématique puissant pour analyser les files d'attente et peuvent fournir des prédictions précises sur le comportement du système sous différentes conditions de charge et de configuration.

1.2 Files d'attente

Une file d'attente est la donnée d'une (ou plusieurs) unité de service où arrivent des clients qui demandent une certaine durée d'utilisation de cette unité (le service demandé par les clients). Quand les clients ne peuvent accéder à cette unité de service, ils patientent dans une file d'attente en attendant d'être servis. La file d'attente peut éventuellement n'accepter qu'un nombre fini de clients, dans ce cas les clients trouvant la file pleine à leur arrivée sont rejetés par le système. Un client peut être servi pendant une certaine période puis abandonné par le serveur. Le service résiduel d'un client est la durée du service qui reste à effectuer, quand celui-ci est nul le client quitte la file d'attente. La charge de la file d'attente est la somme de tous les services résiduels de tous les clients présents.

Le formalisme des files d'attente et les résultats associés ont été introduits puis se sont développés comme une discipline très pratique portant le nom "théorie des files d'attente" dont l'objectif était de construire des modèles permettant de prédire le comportement des systèmes qui fournissent un service à des demandes aléatoires.

En mathématique, une file d'attente est caractérisée de la manière suivante :

- Le processus d'arrivée des clients.
- Le temps de service.
- La capacité du système.
- La discipline de service.

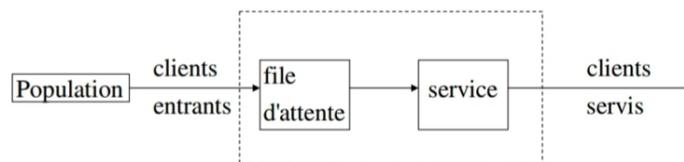


FIGURE 1.1 – Système de file d'attente

1.3 Réseau de files d'attente

Un réseau de files d'attente est un ensemble de files d'attente inter-connectées. Le séjour d'un client dans un réseau consiste à parcourir une partie ou l'ensemble de toutes ces files d'attente, selon un itinéraire déterministe ou stochastique. Les réseaux de files d'attente peuvent être classifiés en deux catégories :

- Les réseaux de files d'attente mono-classe : dans lesquels circule une seule classe de clients. Il y'a deux types réseaux fermés et réseaux ouverts.

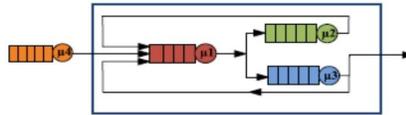


FIGURE 1.2 – Réseau mono-classe ouvert

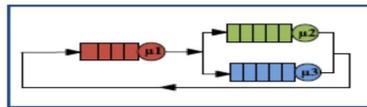


FIGURE 1.3 – Réseau mono-classe fermé

- Les réseaux de files d'attente multi-classes : dans lesquels circulent plusieurs classes de clients. Ces dernières peuvent se distinguer par un schéma de routage spécifique et par des comportements différents au niveau de chaque station, tant au niveau du service que l'ordonnement dans la file d'attente.

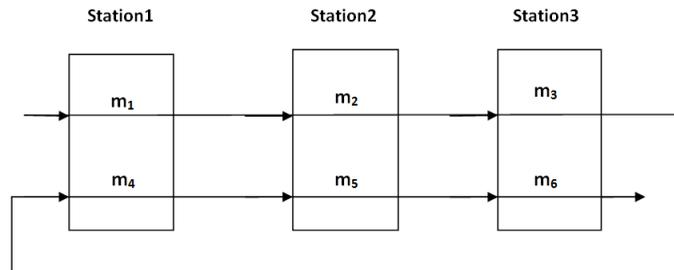


FIGURE 1.4 – Réseau fluide à trois stations et 6 classes de clients

1.3.1 Différentes disciplines de service

Dans cette partie, nous nous concentrons sur certaines disciplines populaires HL (head of the line)(au sein de chaque classe, les clients sont servis dans l'ordre (first in first out)(FIFO)), et nous considérons χ comme l'espace d'état résultant (voir bramson [4] section 4.1).

Nous verrons que l'espace d'état des réseaux de files d'attente spécifiques HL peut être simplifié en supprimant des informations redondantes. De plus, une restriction de temps d'interarrivées et le temps de service permet de supprimer les coordonnées u et v des états, (voir [4]).

- **Disciplines de service FIFO**

Dans les réseaux de files d'attente sous la discipline de service FIFO, les clients sont servis selon le temps passé dans la file d'attente, où les clients les plus anciens sont les premiers dans le service.

Soit $N_j(t) = \sum_{k \in C(j)} Q_k(t)$ la longueur de la file d'attente à la station j au temps t . Les clients à la station j sont ordonnés selon leur temps d'arrivée (premier entré-premier sorti), c'est à dire, pour $N_j(t) > 0$ nous considérons la suite

$$k_j(t) = (k_{j,1}, k_{j,2}, \dots, k_{j,N_j(t)}),$$

où $k_{j,i}$ désigne le numéro du $i^{\text{ème}}$ client à la station j . Si $N_j(t) = 0$, la suite $k_j(t) = 0$.

Par conséquent, l'âge des clients peut être retiré de la description de l'espace d'état. De plus, comme le réseau de files d'attente est HL, la coordonnée z dans (3.30) peut être également éliminée et l'évolution du processus de Markov peut être décrite par

$$X(t) = ((k_1(t), \dots, k_J(t)), U(t), V_{k_{j,1}(t)}(t), j = 1, \dots, J).$$

Ainsi, l'espace d'état satisfait

$$\chi \subset (\mathbb{Z}_K^\infty)^J \times \mathbb{R}_+^{|\epsilon|+J},$$

où \mathbb{Z}_K^∞ désigne l'ensemble des suites qui prend ces valeurs dans $\mathbb{Z}_K := \{1, 2, \dots, K\}$.

- **Disciplines de service avec priorité**

Sous des disciplines de service avec priorité chaque station ordonne les classes de clients et les sert par priorité, c'est à dire, pour chaque station j , on a l'application $\pi_j : C(j) \rightarrow \{1, \dots, |C(j)|\}$, et la classe \hat{k} avec $\pi_j(\hat{k}) = 1$ est dite une classe de priorité supérieure.

Si la station reçoit un nouveau client, le serveur sert le client de la classe de priorité supérieure de la file d'attente. Ainsi, au sein de chaque classe, les clients sont servis selon l'ordre (FIFO).

Si la file d'attente est vide, la station tourne au ralenti. Les disciplines de service prioritaires peuvent être distinguées en deux sous-disciplines.

- **Disciplines de service avec priorité et préemption**

Si un client d'une classe prioritaire rentre dans la file d'attente, alors le service du premier client est arrêté immédiatement, et le client de la classe prioritaire est servi.

Le serveur revient à son travail sauf s'il n'y a aucun autre client d'une classe prioritaire dans la file d'attente.

En suivant les mêmes arguments que dans le cas (FIFO) le processus de Markov peut être décrit par

$$X(t) = ((Q_1(t), \dots, Q_K(t)), U(t), V(t)),$$

et l'espace d'état peut être considéré comme

$$\chi \subset (\mathbb{Z}_+^K) \times \mathbb{R}_+^{|\epsilon|+K}.$$

- **Disciplines de service avec priorité et non-préemption**

Dans ce cas, même s'il arrive un client d'une classe de priorité plus élevée, la station doit terminer le service du client actuel. Ainsi, à chaque station j il y'a au plus un client servi à l'instant t . Alors, on pose $k_j(t)$ la classe du client servie à l'instant t , dans ce cas le processus de Markov peut être décrit par

$$X(t) = ((k_1(t), \dots, k_J(t)), U(t), V_{k_{j,1}(t)}(t), j = 1, \dots, J).$$

S'il n'y a pas de clients présents à la station j au temps t , alors $k_j(t)$ tend vers zéro et $V_{k_{j,1}(t)}(t) = 0$.

Par conséquent, le processus de Markov est complètement décrit par des éléments de l'espace d'état

$$\chi \subset (\mathbb{Z}_+^K) \times \mathbb{R}_+^{|\epsilon|+K} \times \mathbb{Z}_K^J.$$

1.3.2 Processus de Markov sous-jacent

Dans cette partie, nous présentons le processus stochastique qui décrit l'évolution d'un réseau de files d'attente multi-classes.

Premièrement, nous décrivons l'évolution du processus $X = \{X(t) = (Y(t), Z(t)) \in \chi; t \geq 0\}$ entre les arrivées et les départs des clients, où le processus se développe selon les taux de service constants par morceaux $Z(t)$.

On notera que, sur la base de la discipline de service, la relation entre $Z(t)$ et $Y(t)$ peut être décrite par une fonction mesurable.

Deuxièmement, pour décrire la progression du processus de Markov, quand il y'a une arrivée ou bien un départ des clients dans le réseau, nous considérons les incréments primitifs $\{a_k(n), s_k(n), \phi^k(n) : k \in \{1, 2, \dots, K\}, n \geq 1\}$ pour construire l'évolution du X . Pour le temps t entre les arrivées et les départs, le taux de diminution des temps de services résiduels $V(t) = \nu$ est donné par $Z(t)$.

- Il y'a deux possibilités qui peuvent se produire :

1. L'achèvement du service : $V_k(t^-) = 0$ pour $k \in \{1, 2, \dots, K\}$. La transition du client le plus ancien de classe k est donnée par ϕ^k . Par conséquent, nous avons $V_k(t) = s_k(i)$ ou $V_k(t) = 0$ si $q_k(t^-) > 0$ ou $q_k(t^-) = 0$ respectivement. Le client soit quitte le réseau ou il passe à une autre classe, $W_i(t) = w_i$ augmente avec un taux 1. En outre, les composantes des taux

d'arrivées résiduels $U(t) = u$ diminuent de taux égal à 1 vers 0.

- 2. Nouvelle arrivée :** $U_k(t^-) = 0$ pour $k \in \varepsilon$. Dans ce cas, un couple $(k, 0)$ est ajouté à l'état $Y(t)$ et le temps d'interarrivées résiduels est donné par $U_k(t) = a_k(i)$, où "i" représente le nombre de prochains nouveaux clients de classe k au temps t . Si le client de la classe k arrive à la file d'attente vide, c'est à dire $q_k(t^-) = 0$ alors $V_k(t^-) = 0$. Ainsi, nous avons $V_k(t) = s_k(i)$, où "i" est l'indice du premier temps de service inutilisé à l'instant t .

Théorème 1.3.1. [7] *Le processus stochastique X est déterministe par morceau dont l'espace d'état est χ . En particulier X est un processus de Borel à droite et satisfait la propriété de Markov forte.*

1.3.3 Equations des réseaux de files d'attente

Dans cette partie, nous présentons en détail les équations qui décrivent l'évolution des processus de performances des réseaux de files d'attente multi-classes. En se basant sur les incréments primitifs (a, s, ϕ) , nous définissons les cumulatifs primitifs (E, S, R) du réseau de files d'attente.

Par convention, nous supposons que $E(0) = S(0) = R(0) = 0$.

- Les suites $a_1, a_2, \dots, a_K, s_1, s_2, \dots, s_K$ et $\phi^1, \phi^2, \dots, \phi^K$ sont indépendantes identiquement distribuées et mutuellement indépendantes.
- Les premiers moments satisfaisant

$$\alpha_k = \mathbb{E}[a_k(1)]^{-1} < \infty, \quad \text{pour } k \in \varepsilon,$$

$$\mu_k = \mathbb{E}[s_k(1)]^{-1} < \infty, \quad \text{pour } k \in \{1, 2, \dots, K\},$$

$$P_k = \mathbb{E}[\Phi^k(1)]^{-1} \geq 0, \quad \text{pour } k \in \{1, 2, \dots, K\},$$

- $E(t)$, processus d'arrivée externe.
- Le temps d'arrivée du $n^{\text{ième}}$ client de la classe k est donné par $a_k(1), \dots, a_k(n)$.
- Le processus $E_k(t)$ compte les clients de classe k arrivant de l'extérieur au temps t , il est défini par

$$E_k(t) = \max\{n \in Z_+ : a_k(1) + \dots + a_k(n) \leq t\}.$$

- $S(t)$, processus de service cumulatif.
- La composante $S_k(t)$ compte les achevements de service des clients de classe k dans la période de temps $[0; t]$, dans ce cas la station attribue toute sa capacité à cette classe de clients, c'est

à dire

$$S_k(t) = \max\{n \in \mathbb{Z}_+ : s_k(1) + \dots + s_k(n) \leq t\}.$$

- $R(n)$, processus de routage.

Pour chaque $n \in \mathbb{Z}_+$ et pour chaque classe de clients $k \in \{1, 2, \dots, K\}$ le processus de routage est défini par

$$R^k(n) = \sum_{i=1}^n \phi^k(i)$$

Bien entendu, les processus E , S et R sont càdlàg (continues à droite ayant une limite à gauche).

Avant de définir les équations qui décrivent l'évolution du réseau de files d'attente, nous devons introduire le processus d'allocation, noté par $T = \{T(t), t \geq 0\}$. Ce processus présente le temps consacré dans une station pour servir les classes de clients présentes à cette dernière.

Afin d'être précis, $T_k(t)$ désigne le temps de service accumulé $c(k)$ consacré à la station pour servir les clients de classe k dans l'intervalle $[0, t]$.

Le processus d'allocation est déterminé par une discipline de service. De la définition il s'ensuit que $T(\cdot)$ est croissant.

Etant donné le processus d'allocation T , le nombre de clients qui quittent la classe k à l'instant t est donné par $S_k(T_k(t))$.

De plus, le nombre de clients de classe l qui ont passé de la classe l vers la classe k dans la période $[0, t]$ est donné par $R_k^l(S_l(T_l(t)))$.

Ainsi, le niveau de fluide de la classe k au temps t peut être décrit par l'équation d'équilibre suivante

$$Q_k(t) = Q_k(0) + E_k(t) + \sum_{l=1}^K R_k^l(S_l(T_l(t))) - S_k(T_k(t)),$$

où $Q_k(0)$ désigne le nombre de clients qui se trouvent dans la file d'attente au temps $t = 0$. Les arrivées totales de clients de classe k à l'instant t sont définies par

$$A_k(t) = E_k(t) + \sum_{l=1}^K R_k^l(S_l(T_l(t))),$$

et les départs dans $[0, t]$ sont

$$D_k(t) = S_k(T_k(t)).$$

L'équation d'équilibre ci-dessus peut être écrite sous la forme simple suivante

$$Q_k(t) = Q_k(0) + A_k(t) - D_k(t).$$

En outre, il existe des processus additionnels qui sont utilisés pour décrire l'évolution du réseau de files d'attente multi-classes.

Le processus $W = \{W(t), t \geq 0\}$ est appelé processus de la charge de travail immédiate.

La composante $W_j(t)$ reflète le temps résiduel nécessaire pour servir tous les clients qui sont actuellement en attente d'être servis à la station j . En outre, pour $n = (n_1, \dots, n_k)^T$, soit $\Gamma(n) = (\Gamma_1(n_1), \dots, \Gamma_K(n_K))^T$ le service cumulatif défini par

$$\Gamma_k(n_k) = \sum_{i=1}^{n_k} s_k(i).$$

Ainsi, la charge de travail immédiate peut être caractérisée par

$$W(t) = C\Gamma(Q(0) + A(t)) - CT(t),$$

où C désigne la matrice composition.

En outre, le processus $I = \{I(t), t \geq 0\}$ de dimension J est appelé processus de temps de repos, ce processus désigne le temps total pendant lequel les stations ne fonctionnaient pas dans la période de temps $[0, t]$. Le processus de temps de repos peut être décrit par la condition suivante

$$I(t) = et - CT(t),$$

avec $e = (1, \dots, 1)^T$. Comme $T(\cdot)$ est croissant, il s'ensuit que $I(\cdot)$ est également croissant.

En se basant sur ces processus on trouve une autre propriété essentielle sur les réseaux de files d'attente multi-classes. Cette propriété est appelée la propriété "non-idling".

Cela signifie qu'une station prend un temps de repose à l'instant t que si la file d'attente est vide, ou encore une station sert une classe d'un client prioritaire dès qu'elle est prête à être servie et qui ne peut pas être retardée si le serveur n'a rien d'autre à faire, on dit dans ce cas que l'ordonnancement est non oisif c'est à dire qu'il fonctionne sans insertion de temps creux (non-idling ou work conserving).

Cela signifie que, si pour la station j , il s'avère que $I_j(t_1) < I_j(t_2)$ pour $t_1 < t_2$, alors il existe $t \in [t_1, t_2]$ tel que $W_j(t) = 0$. Comme $I(\cdot)$ est continu, la propriété non-idling peut s'écrire aussi comme

$$\int_0^{\infty} W_j(t) dI_j(t) = 0.$$

Finalement, avant de donner les équations des réseaux de files d'attente multi-classes, nous notons que la propriété HL d'un réseau de files d'attente peut être exprimée comme suit

$$\Gamma(D(t)) \leq T(t) < \Gamma(D(t) + e),$$

où les inégalités doivent être comprises par composante. Sous forme vectorielle, les équations d'un réseau de files d'attente multi-classes peuvent être résumées par

$$Q_k(t) = Q_k(0) + A_k(t) - D_k(t), \quad (1.1)$$

$$T(\cdot) \text{ est croissant, avec } T(0) = 0, \quad (1.2)$$

$$W(t) = C\Gamma(Q(0) + A(t)) - CT(t), \quad (1.3)$$

$$I(t) = et - CT(t), \quad I(\cdot) \text{ est croissant,} \quad (1.4)$$

$$I_j(t) \text{ augmente seulement si } W_j(t) = 0, \quad j \in \{1, \dots, J\}, \quad (1.5)$$

$$\text{Les conditions additionnelles sur } (Q(\cdot), T(\cdot)) \text{ sont spécifiques à la discipline.} \quad (1.6)$$

Dans ce qui suit nous nous référons à (1.1)-(1.6) comme des équations des réseaux de files d'attente.

Les processus définissant les équations de base de réseaux de files d'attente déterminent l'évolution de ce dernier. Pour cette raison, nous appelons

$$X(t) = (A(t), D(t), T(t), W(t), I(t), Q(t)),$$

le processus de réseau de files d'attente.

Stabilité des modèles stochastiques

2.1 Heuristique

Une grande partie des résultats de probabilité obtenus dans la section (2.2) est basée sur la méthode de Lyapunov. Ce sont les travaux du début du siècle qui ont permis d'obtenir une méthode d'étude de la stabilité de l'ordre mécanique dans le cadre de systèmes régis par des équations différentielles. Le concept est basé sur le développement de fonctions connues sous le nom de Lyapunov. Cette première section présente les résultats obtenus.

La technique a ensuite été modifiée pour examiner la stabilité des modèles stochastiques. La deuxième section examine leurs résultats généraux.

On soulève maintenant l'étude de la stabilité et de la dynamique des systèmes mécaniques. Ce problème conduit mathématiquement à l'étude de la limite de comportement des solutions aux équations différentielles de la forme

$$\dot{x} = f(x), \quad (2.1)$$

où f est une fonction continuellement différentiable définie sur un W ouvert de \mathbb{R}^n à des valeurs dans \mathbb{R}^n .

En premier lieu, la stabilité du système est définie par rapport aux perturbations de données initiales :

Définition 2.1. *Une solution $x(t)$ d'une équation différentielle est considérée comme stable si, pour toute valeur de $\varepsilon > 0$, il existe une valeur de $\delta > 0$ telle que toute autre solution $y(t)$ satisfaisant $\|y(0) - x(0)\| \leq \delta$, on ait $\|y(t) - x(t)\| \leq \varepsilon$, pour tout $t \geq 0$. De plus, si $(y(t) - x(t))$ tend vers 0 lorsque t tend vers $+\infty$, on dit que $x(t)$ est stable asymptotiquement.*

Quand nous savons intégrer les équations différentielles, ce problème, bien sûr, ne présente pas de difficultés. Mais il est important d'avoir des méthodes qui permettent de le résoudre, indépendamment de la possibilité de cette intégration. Avant Lyapunov, les problèmes de stabilité étaient habituellement résolus en linéarisant les équations différentielles et en négligeant tout ce qui était d'ordre supérieur. Ce procédé entraînait une simplification importante, surtout lorsque les coefficients de ces équations différentielles sont constants. Certains comme, Routh par exemple, reconnaissant ce procédé peu rigoureux, ne se bornèrent pas à une première approximation, mais

utilisèrent des approximations d'ordre deux et même plus. En opérant ainsi, on avance peu, car, en général, par cette voie, on obtient seulement une représentation fiable des fonctions dans les limites d'un intervalle de temps compact; ce qui assurément, ne donne pas de nouveaux éléments pour obtenir des conclusions quelconques sur la stabilité.

Lyapunov a appelé point d'équilibre de (2.1) tout point \tilde{x} de W vérifiant $f(\tilde{x}) = 0$. Il est clair que la fonction constante $x(t) \equiv \tilde{x}$ est solution de l'équation différentielle.

Lyapunov trouvera un critère de stabilité en utilisant des fonctions V positives définies sur un voisinage de \tilde{x} et décroissantes sur le chemin déterminé par f . Voici son résultat principal :

Théorème 2.1.1. *Soit $\tilde{x} \in W$ un point d'équilibre pour (2.1). Soit $V : U \rightarrow \mathbb{R}^+$ une fonction continue définie sur un voisinage U de \tilde{x} , différentiable sur $U \setminus \tilde{x}$; telle que*

$$V(\tilde{x}) = 0 \text{ et } V(x) > 0 \text{ si } x \neq \tilde{x}.$$

$$\text{Si } \frac{dV}{dx}(f(x)) \leq 0 \text{ sur } U \setminus \tilde{x},$$

de plus, \tilde{x} est stable, si

$$\frac{dV}{dx}(f(x)) < 0 \text{ sur } U \setminus \tilde{x},$$

le point \tilde{x} est asymptotiquement stable. La fonction V est appelée fonction de Lyapunov.

La preuve de ce théorème peut être trouvée dans [23]. La méthode décrite est puissante, à condition que les fonctions de Lyapunov puissent être trouvées. Bien que ce problème ait été abordé par de nombreux mathématiciens et qu'un certain nombre de propriétés aient été démontrées, il n'existe actuellement aucune méthode générale permettant de trouver les fonctions de Lyapunov. Il existe parfois des fonctions naturelles à essayer. Dans le cas de systèmes mécaniques ou électriques, l'énergie est souvent une fonction de Lyapunov.

2.2 Résultat généraux

La méthode utilisée par Lyapunov a été reprise dans un cadre probabiliste. De nouvelles fonctions de Lyapunov permettent d'établir dans le théorème 2 de [15], des conditions de stabilité d'une chaîne de Markov.

Le modèle dit stable si la chaîne de Markov correspondante est récurrente, c'est à dire que partant de n'importe quel état, le retour à une position donnée est certain.

La chaîne sera dite récurrente positive si les variables indiquant le temps de retour à une position donnée sont intégrables. S'il est également apériodique, il est ergodique.

Nous donnerons ensuite, dans la proposition (2.2.3), des conditions d'existence des moments d'ordre α ($\alpha \geq 1$) de variables aléatoires indiquant le temps de retour d'une chaîne de Markov, soit vers un point, soit vers un sous-ensemble de l'espace d'état.

Dans cette section, on se place sur un espace d'états dénombrable S et on considère une chaîne de Markov $(M_n)_{n \geq 0}$ homogène, irréductible, apériodique à valeurs dans S . On note $P = (p(x, y))_{(x, y)}$

la matrice de transition, \mathcal{F}_n la tribu engendrée par M_0, \dots, M_n et on note $\mathbb{E}_x(\cdot)$ l'espérance d'une variable lorsque la chaîne est en x à l'instant initial. D'autres résultats ont été obtenus à l'aide des fonctions de Lyapunov. Citons, en particulier, les algorithmes stochastiques (voir [12]), la stabilité des solutions d'équations différentielles stochastiques (voir [16]).

Les résultats indiqués dans cette section se trouvent dans le livre de Meyn et Tweedie [18] et Menshikov [14].

Définition 2.2. Si τ est un temps d'arrêt relativement à la filtration \mathcal{F} la suite induite associée à τ est une suite croissante t_n définie par

$$\begin{aligned} t_0 &= 0, \\ t_n &= t_{n-1} + \tau \circ \theta^{t_{n-1}}, \end{aligned}$$

pour $n \geq 1$. La chaîne induite est la suite (M_{t_n}) et la filtration induite \mathcal{F}^τ est la filtration associée, $\mathcal{F}^\tau = (\mathcal{F}_n^\tau) = (\mathcal{F}_{t_n})$.

Dans le cas où la chaîne de Markov est ergodique et stationnaire, la translation θ^1 est l'endomorphisme de l'espace de probabilité. Si τ le temps d'atteinte de chaîne de Markov, l'application θ^τ est l'endomorphisme induit défini (3.7).

2.2.1 Ergodicité d'une chaîne de Markov

Le principe fondamental de l'ergodicité des chaînes de Markov est énoncé dans la proposition suivante.

Proposition 2.2.1.

Si

$$F \text{ est un sous-ensemble fini de } S, \text{ et } T_F = \inf \{k \geq 0 / M_k \in F\},$$

et si

$$g(x) = \mathbb{E}_x(T_F) \text{ est fini, } \forall x \in S, \text{ et } \mathbb{E}_x(g(M_1)) \leq +\infty, \forall x \in F$$

la chaîne de Markov est ergodique.

Démonstration. Comme la variable T_F est finie presque sûrement, la chaîne de Markov passe une infinité de fois dans F .

Pour $p \geq 1$, X_p désigne l'élément de F visité par (M_n) lors de la $p^{\text{ième}}$ visite à F .

La propriété de Markov et l'irréductibilité de (M_n) montrent que (X_n) est aussi une chaîne de Markov irréductible sur l'ensemble fini F .

La chaîne (X_n) a donc une probabilité invariante π_F . Le temps de retour strict T_F^+ à F est défini comme

$$T_F^+ = \inf \{k > 0 / M_k \in F\},$$

et la mesure π sur S par $\pi(f) = \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{T_F^+-1} f(M_i) \right)$, pour toute fonction f bornée sur S .

On va montrer que la mesure π ainsi définie est invariante pour la chaîne de Markov (M_n) . Si x est un élément de S , la définition précédente donne

$$\pi(x) = \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{T_F^+-1} \mathbf{1}_{\{M_i=x\}} \right),$$

d'où

$$\pi(x) \mathbb{E}_{\pi_F}(f(M_1)) = \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{+\infty} \mathbf{1}_{\{M_i=x, 0 \leq i < T_F^+\}} \mathbb{E}_x(f(M_1)) \right);$$

pour $i \in \mathbb{N}$ l'événement $\{0 \leq i < T_F^+, M_i = x\}$ est \mathcal{F}_i -mesurable et en utilisant la propriété de Markov de (M_n) , on a

$$\begin{aligned} \pi(x) \mathbb{E}_x(f(M_1)) &= \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{+\infty} \mathbf{1}_{\{M_i=x, 0 \leq i \leq T_F^+\}} \mathbb{E}_{\pi_F}(f(M_{i+1}) \mid \mathcal{F}_i) \right) \\ &= \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{+\infty} \mathbb{E}_{\pi_F}(\mathbf{1}_{\{M_i=x, 0 \leq i \leq T_F^+\}} f(M_{i+1}) \mid \mathcal{F}_i) \right) \\ &= \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{T_F^+-1} \mathbf{1}_{\{M_i=x\}} f(M_{i+1}) \right). \end{aligned}$$

La mesure π vérifie donc

$$\begin{aligned} \mathbb{E}_{\pi}(f(M_1)) &= \sum_{x \in S} \pi(x) \mathbb{E}_x(f(M_1)) = \mathbb{E}_{\pi_F} \left(\sum_{i=0}^{T_F^+-1} f(M_{i+1}) \right) \\ &= \mathbb{E}_{\pi_F} \left(\sum_{i=1}^{T_F^+-1} f(M_i) \right) + \mathbb{E}_{\pi_F}(f(M_{T_F^+})). \end{aligned}$$

Par définition de la chaîne (X_n) , $M_{T_F^+} = X_1$, et comme π_F est la probabilité invariante de la chaîne (X_n) on déduit

$$\mathbb{E}_{\pi_F}(f(M_{T_F^+})) = \mathbb{E}_{\pi_F}(f(X_1)) = \mathbb{E}_{\pi_F}(f(X_0)) = \mathbb{E}_{\pi_F}(f(M_0)).$$

Par conséquent $\mathbb{E}_{\pi_F}(f(M_1)) = \pi(f)$, π est donc une mesure invariante pour la chaîne (M_n) . Par définition de T_F^+ , en décomposant suivant la première transition de la chaîne de Markov il vient

$$\begin{aligned} \pi(1) &= \mathbb{E}_{\pi_F}(T_F^+) \\ &= 1 + \mathbb{E}_{\pi_F}(\mathbb{E}_{M_1}(T_F^+) \mathbf{1}_{\{M_1 \notin F\}}) \\ &= 1 + \mathbb{E}_{\pi_F}(\mathbb{E}_{M_1}(T_F) \mathbf{1}_{\{M_1 \notin F\}}) \end{aligned} \tag{2.2}$$

d'où l'inégalité

$$\pi(1) \leq 1 + \mathbb{E}_{\pi_F}(g(M_1)) \leq \sup_{x \in F} \mathbb{E}_x(g(M_1)) < +\infty.$$

La chaîne de Markov irréductible (M_n) a une mesure invariante finie, elle est donc ergodique. \square

Le résultat principal concernant l'ergodicité peut être maintenant établi. La formulation est due Filonov [15].

Théorème 2.2.1. *S'il existe une fonction $f : S \rightarrow \mathbb{R}_+, K, \gamma > 0$ et un temps d'arrêt $\tau \geq 1$ intégrable tels que*

$$\mathbb{E}_x(f(M_\tau) - f(x)) \leq -\gamma \mathbb{E}_x(T) \text{ si } f(x) > K,$$

en notant F l'ensemble $\{x/f(x) \leq K\}$, le temps d'atteinte T_F de F ,

$$T_F = \inf \{k \geq 0 / f(M_k) \leq K\} \text{ est intégrable,}$$

et

$$\mathbb{E}_x(T_F) \leq f(x)/\gamma, \quad x \in S.$$

Si de plus l'ensemble F est fini et

$$\mathbb{E}_x(f(M)_1) < +\infty \text{ quand } f(x) \leq K,$$

Donc la chaîne de Markov (M_n) est ergodique.

Une fonction satisfaisant les conditions ci-dessus peut être interprétée comme une fonction énergie de la chaîne de Markov, cette fonction décroît en moyenne au cours du temps quand l'état initial est très grand pour la fonction. Ce type de résultat (fonction de Lyapunov) existe sous de nombreuses formes.

Démonstration. Soit (t_n) la suite induite associée à τ , si $v = \inf \{k \geq 0 / t_k \geq T_F\}$, l'ensemble $\{v < k\} = \{t_k > T_F\}$ est $\mathcal{F}_k^\tau (= \mathcal{F}_{t_k})$ -mesurable puisque T_F est un temps d'arrêt relativement à la filtration (\mathcal{F}_n) .

Par conséquent v est un temps d'arrêt par rapport à la filtration $\mathcal{F}^\tau = (\mathcal{F}_k^\tau)$.

Posons $X_n = f(M_{t_n}) + \gamma t_n$. La suite de variables aléatoires positives (X_n) est adaptée à la filtration (\mathcal{F}_n^τ) et

$$\mathbb{E}_x(X_{n+1} | \mathcal{F}_n^\tau) = \mathbb{E}_x(f(M_{t_n} + \tau(\theta_{t_n})) + \gamma(t_n + \tau(\theta_{t_n})) | \mathcal{F}_n^\tau),$$

La propriété de Markov forte donne l'identité

$$\mathbb{E}_x(X_{n+1} | \mathcal{F}_n^\tau) = \gamma t_n + \mathbb{E}_{M_{t_n}}(f(M_\tau) + \gamma \tau);$$

l'inégalité $f(M_{t_n}) > K$ est vraie sur l'ensemble $\{v > n\} = \{t_n < T_F\}$, par conséquent

$$\mathbb{E}_x(X_{n+1} | \mathcal{F}_n^\tau) - X_n = \mathbb{E}_{M_{t_n}}(f(M_\tau) - f(M_{t_n}) + \gamma \tau) \leq 0,$$

sur l'ensemble $\{v > n\}$ selon la condition 1. En d'autres termes, $(X_{v \wedge n})$ est une sur-martingale positive par rapport à la filtration \mathcal{G} . En particulier, $\mathbb{E}_x(X_{v \wedge n}) \leq X_0$, Soit

$$\mathbb{E}_x(f(M_{t_{v \wedge n}}) + \gamma t_{v \wedge n}) \leq f(x).$$

La fonction f étant positive, le théorème de convergence monotone montre l'inégalité

$$\mathbb{E}_x(t_v) \leq f(x)/\gamma.$$

Comme $t_v \geq T_F$, on en déduit la majoration du temps de retour à F , $\mathbb{E}_x(T_F) \leq f(x)/\gamma$ pour tout $x \in S$, et pour tout $x \in F$, on a l'inégalité

$$\mathbb{E}_x(\mathbb{E}_{M_1}(T_F)) \leq \mathbb{E}_x\left(\frac{f(M_1)}{\gamma}\right) < +\infty.$$

D'après la proposition (2.2.1), on conclure que la chaîne de Markov est ergodique. \square

2.2.2 Transience d'une chaîne de Markov

La proposition suivante donne un critère simple. C'est une variante d'un résultat plus général dû à Lamperti [17].

Proposition 2.2.2. *S'il existe une fonction $f : S \rightarrow \mathbb{R}_+$ et $K, \gamma > 0$ tels que $\sup \{f(x), x \in S\} \geq K$ et*

1. $\mathbb{E}_x(f(M_1) - f(x)) \geq \gamma$ si $f(x) \geq K$,

2. $\sup_{x \in S} \mathbb{E}_x(|f(M_1) - f(x)|^2) < +\infty$,

alors la chaîne de Markov (M_n) est transiente.

Cette proposition présente des hypothèses qui sont généralement faciles à vérifier, en particulier dans le contexte des files d'attente.

2.2.3 Propriétés sur les temps de retour d'une chaîne de Markov

Les fonctions de Lyapunov permettent d'étudier les moments des variables indiquant le temps de retour à un état donné selon des hypothèses de base. Considérons l'élément 0 de l'espace d'état S et F un sous ensemble fini de S contenant 0, ainsi que les variables aléatoires T_0, T_F et T_F^+ définies comme suit :

$$T_0 = \inf \{n \geq 0 / M_n = 0\},$$

$$T_F = \inf \{n \geq 0 / M_n \in F\},$$

$$T_F^+ = \inf \{n \geq 1 / M_n \in F\}.$$

Proposition 2.2.3. *S'il existe une fonction f positive définie sur S , un réel $\alpha > 1$ et des réels strictement positifs K, γ et un sous-ensemble F de S vérifiant*

$$\mathbb{E}_x(f(M_1) - f(M_0)) \leq -\gamma \text{ sur } S, \tag{2.3}$$

$$\mathbb{E}_x(|f(M_1) - f(M_0)|^\alpha) \leq K, \tag{2.4}$$

alors

$$\mathbb{E}(T_F^\alpha | M_0) \leq A + B[f(M_0)]^\alpha, \tag{2.5}$$

où A et B sont des constantes positives ne dépendant que de α, γ et K . Si de plus F est fini, alors

$$\forall y \in F, \mathbb{E}_y(T_0^\alpha) < +\infty.$$

Remarque 2. *On peut, avec des hypothèses comparables, établir des propriétés similaires, encore à l'aide de fonctions de Lyapunov pour d'autres moments et en particulier des moments exponentiels.*

Démonstration. Les idées de la première partie de la démonstration sont dues à Malyshev et Menshikov. La variable $f(M_1) - f(M_0)$ ayant un moment d'ordre $\alpha > 1$ majoré indépendamment de M_0 , on peut trouver un réel C supérieur à 1 et $\gamma/2$ tels que la variable

$$\Gamma = (f(M_1) - f(M_0))\mathbb{1}_{\{f(M_1) - f(M_0) > -c\}} + \frac{\gamma}{2}$$

vérifie

$$\mathbb{E}(\Gamma) \leq \frac{-\gamma}{4} \text{ sur } (T_F > 0).$$

Posons

$$\begin{aligned} Y_n &= f(M_n) + n\gamma/2, \quad \Delta_n = Y_n - Y_{n-1}, \\ X_n &= (1 + \delta Y_n)^\alpha \mathbf{1}_{\{T_F \geq n\}}. \end{aligned}$$

Le réel δ appartient à l'intervalle $[0, 1/C]$. Sur $(T_F \geq n)$ on a

$$\begin{aligned} \mathbb{E}(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) &= X_{n-1} \mathbb{E} \left(\left(1 + \frac{\delta \Delta_n}{1 + \delta Y_{n-1}} \right)^\alpha - 1 \mid \mathcal{F}_{n-1} \right), \\ &\leq g X_{n-1} \left(\frac{\delta}{1 + \delta Y_{n-1}}; M_{n-1} \right), \end{aligned}$$

où g est la fonction définie sur $[0, 1/C] \times S$ par

$$g(\beta; x) = \mathbb{E}_x((1 + \beta\Gamma)^\alpha - 1).$$

Cette application est dérivable par rapport à β . En effet,

$$\begin{aligned} \left| \frac{\partial(1 + \beta\Gamma)^\alpha}{\partial\beta} \right| &= |\alpha\Gamma(1 + \beta\Gamma)^{\alpha-1}|, \\ &\leq \alpha(1 + |\Gamma|)^\alpha. \end{aligned}$$

Il suffit alors d'utiliser le théorème de convergence dominée. De plus,

$$\begin{aligned} \left| \frac{\partial g}{\partial\beta}(\beta; x) - \frac{\partial g}{\partial\beta}(0; x) \right| &\leq \int_{\Gamma \leq 1} \alpha |\Gamma((1 + \beta\Gamma)^{\alpha-1} - 1)\rho| dP \\ &\quad + \int_{\Gamma > 1} \alpha \Gamma^\alpha \left[\left(\frac{1}{\Gamma} + \beta \right)^{\alpha-1} - \left(\frac{1}{\Gamma} \right)^{\alpha-1} \right] dP. \end{aligned}$$

Les deux termes de droite tendent vers 0 uniformément en x lorsque β tend vers 0. Le premier est majoré par $\alpha C \max[(1 - \beta C)^{\alpha-1} - 1; (1 + \beta)^{\alpha-1} - 1]$.

Pour le second, on utilise l'uniforme continuité de la fonction $t^{\alpha-1}$ sur $[0, 1]$.

$\partial g/\partial\beta$ est ainsi continue en 0 uniformément par rapport à x . Nous avons

$$g(0, x) = 0$$

et

$$\frac{\partial g(0, x)}{\partial\beta} \leq -\frac{\gamma\alpha}{4} \text{ sur } \{T_F > 0\}.$$

Il existe alors un réel β_0 strictement positif tel que $g(\beta, x) \leq 0$ pour tout $\beta \leq \beta_0$ et $x \notin F$.

En choisissant $\delta \leq \beta_0$, on obtient

$$\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1} \text{ sur } (T_F \leq n),$$

d'autre part,

$$\mathbb{E}(X_n | \mathcal{F}_{n-1}) = 0 \leq X_{n-1} \text{ sur } (T_F \leq n).$$

X_n est une sur-martingale et T_F étant un temps d'arrêt,

$$\left(\frac{\gamma\delta}{2}\right)^\alpha \mathbb{E}_x((T_F \wedge n)^\alpha) \leq \mathbb{E}_x(X_{T_F \wedge n}) \leq X_0.$$

En passant à la limite

$$\mathbb{E}_x(T_F^\alpha) \leq \left(\frac{2 + 2\delta f(x)}{\gamma\delta}\right)^\alpha.$$

En utilisant pour $p = 2$ l'inégalité

$$\left(\sum_{i=1}^p a_i\right)^\alpha \leq p^{\alpha-1} \sum_{i=1}^p a_i^\alpha, \quad (2.6)$$

on démontre (2.5).

Remarquons que si y est un élément de F , alors

$$\mathbb{E}_y[(T_F^+)^\alpha] < +\infty.$$

En effet,

$$\begin{aligned} \mathbb{E}_y[(T_F^+)^\alpha] &= \sum_{x \in S} p(y, x) [\mathbb{E}_x(T_F + 1)^\alpha] \\ &\leq 2^{\alpha-1} \sum_{x \in S} p(y, x) [\mathbb{E}_x(T_F^\alpha) + 1] \\ &\leq 2^{\alpha-1} \left[(A + 1) + B \sum_{x \in S} p(y, x) f^\alpha(x) \right] \\ &\leq 2^{\alpha-1} [(A + 1) + B \mathbb{E}_y(f^\alpha(M_1))] < +\infty. \end{aligned}$$

On peut ainsi définir une chaîne de Markov (X_n) induite par la trace de (M_n) dans F . La probabilité de transition entre deux éléments y et \dot{y} de F est

$$P^F(y, \dot{y}) = P_y(M_{T_F} = \dot{y}).$$

Cette chaîne est irréductible, finie, donc récurrente positive. Pour y élément de F , on note C_y l'ensemble des chemins dans F d'origine y et tronqués dès le premier passage en 0. La longueur de ces chemins est presque sûrement finie. Un tel chemin sera noté $C = (y_0, y_1, \dots, y_k)$ où $y_0 = y$ et $y_k = 0$. Partant de y , on obtient

$$\mathbb{E}_y(T_0^\alpha) = \sum_{C=(y, y_1, \dots, y_{k-1}, 0) \in C_y} P^F(C) (\mathbb{E}(T_0^\alpha | C))$$

et en utilisant la formule (2.6),

$$\mathbb{E}(T_0^\alpha | C) \leq k^{\alpha-1} \sum_{i=1}^{k-1} \mathbb{E}_{y_i}(T_F^\alpha | M_{T_F} = y_{i+1}) \leq k^\alpha \Theta,$$

où

$$\Theta = \max \left\{ \mathbb{E}_y[(T_F^+)^{\alpha} | M_{T_F^+}^+ = y] / P^F(y, y) > 0 \right\}.$$

Il en résulte

$$\mathbb{E}_y(T_0^{\alpha}) \leq \Theta \mathbb{E}_y^F(T_0^{\alpha}) < +\infty,$$

où T_0 est le temps de retour à 0 de la chaîne (X_n) induite sur l'espace d'état fini F . On peut conclure d'après (2.5). \square

Pour terminer cette partie, nous donnons un résultat asymptotique déjà montré par Malyshev et Menshikov sous une forme différente [13]. Cette propriété est applicable en particulier aux suites de chaînes de Markov vérifiant les hypothèses de la proposition (2.2.3) de façon uniforme. Nous remarquons que l'hypothèse $\alpha \geq 1$ n'est pas suffisante pour la convergence des mesures invariantes.

Proposition 2.2.4. *On considère une suite de chaînes de Markov (X_n^N) sur un espace d'état S dénombrable de probabilités de transition P^N et une chaîne (X_n) de probabilités de transition P sur le même espace vérifiant*

$$\forall (x; y) \in S^2, \lim_{N \rightarrow +\infty} P^N(x; y) = P(x; y),$$

$$\exists \alpha > 1 \text{ tel que } J = \sup_N \mathbb{E}_0^N[(T_0^+)^{\alpha}] < +\infty,$$

alors,

$$\mathbb{E}_0[(T_0^+)^{\alpha}] \leq J \tag{2.7}$$

et les mesures invariantes π_N convergent en loi vers la mesure invariante π .

Démonstration. Soit k un entier positif. Si S^k est l'ensemble des chemins de S de longueur K , pour tout $\varepsilon > 0$, on peut extraire une partie finie Θ de S^k telle que, pour N assez grand, $P_0(\Theta)$ et $P_0^N(\Theta)$ soient supérieurs à $(1 - \varepsilon)$. L'ensemble $\{(T_0^+ \leq k)\} \cap \Theta$ étant fini, pour N assez grand,

$$|P^N[(T_0^+)^{\alpha} \leq k^{\alpha} \cap \Theta] - P[(T_0^+)^{\alpha} \leq k^{\alpha} \cap \Theta]| \leq \varepsilon.$$

On en déduit alors

$$\lim_{N \rightarrow +\infty} P^N[(T_0^+)^{\alpha} \leq k^{\alpha}] = P[(T_0^+)^{\alpha} \leq k^{\alpha}].$$

et d'après le lemme de Fatou,

$$\mathbb{E}_0((T_0^+)^{\alpha}) = \sum_{k=1}^{+\infty} P_0[(T_0^+)^{\alpha} < k^{\alpha}] \leq \liminf \sum_{k=1}^{+\infty} P_0^N[(T_0^+)^{\alpha} > k^{\alpha}] \leq J.$$

Ce qui prouve (2.7).

Pour la convergence des mesures invariantes, montrons tout d'abord que pour toute fonction f bornée

$$f : S \rightarrow \mathbb{R},$$

on a

$$\lim_{N \rightarrow +\infty} \mathbb{E}_0^N \left(\sum_{i=1}^{T_0} f(X_i) \right) = \mathbb{E}_0 \left(\sum_{i=1}^{T_0} f(X_i) \right) < +\infty. \tag{2.8}$$

Si f est positive, pour tout entier naturel k ,

$$\begin{aligned} \mathbb{E}_0 \left(\sum_{i=1}^{T_0^N} f(X_i^N) \right) &= \mathbb{E}_0 \left(\sum_{i=1}^{T_0^N \wedge k} f(X_i^N) \right) + \mathbb{E}_0 \left(\sum_{i=1}^{T_0^N} f(X_i^N) \mathbf{1}_{(T_0^N > k)} \right). \\ \mathbb{E}_0 \left(\sum_{i=1}^{T_0^N} f(X_i^N) \mathbf{1}_{(T_0^N > k)} \right) &\leq \|f\| \mathbb{E}_0(T_0^N \mathbf{1}_{(T_0^N > k)}) \\ &\leq \frac{\|f\|}{k^{\alpha-1}} \mathbb{E}_0[(T_0^N)^\alpha] \\ &\leq \frac{J \|f\|}{k^{\alpha-1}}. \end{aligned}$$

Par ailleurs, grâce à la convergence des probabilités de transition,

$$(X_1^N, \dots, X_k^N) \xrightarrow{\mathcal{L}} (X_1, \dots, X_k),$$

doù

$$\lim_{N \rightarrow +\infty} \mathbb{E}_0 \left(\sum_{i=1}^{T_0^N \wedge k} f(X_i^N) \right) = \mathbb{E}_0 \left(\sum_{i=1}^{T_0 \wedge k} f(X_i) \right). \quad (2.9)$$

On remarque que le terme de gauche est majoré pour tout k par $J \|f\|$, donc la suite

$$k \rightarrow \mathbb{E}_0 \left(\sum_{i=1}^{T_0 \wedge k} f(X_i) \right)$$

est bornée. f étant positive, cette suite converge vers $\mathbb{E}_0 \left(\sum_{i=1}^{T_0} f(X_i) \right)$. Soit $\varepsilon > 0$ fixé. On peut trouver k_0 vérifiant (2.10) et (2.11).

$$\left| \mathbb{E}_0 \left(\sum_{i=1}^{T_0} f(X_i) \right) - \mathbb{E}_0 \left(\sum_{i=1}^{T_0 \wedge k_0} f(X_i) \right) \right| \leq \varepsilon. \quad (2.10)$$

$$\forall N > 0 \quad \mathbb{E}_0 \left(\sum_{i=k_0+1}^{T_0^N} f(X_i^N) \mathbf{1}_{(T_0^N > k_0)} \right) \leq \frac{J \|f\|}{k_0^{\alpha-1}} \leq \varepsilon. \quad (2.11)$$

Daprès (2.9), il existe N_{k_0} vérifiant

$$\forall N \geq N_{k_0} \left| \mathbb{E}_0 \left(\sum_{i=1}^{T_0 \wedge k_0} f(X_i) \right) - \mathbb{E}_0 \left(\sum_{i=1}^{T_0 \wedge k_0} f(X_i^N) \right) \right| \leq \varepsilon. \quad (2.12)$$

Daprès (2.10), (2.11) et (2.12) nous obtenons

$$\left| \mathbb{E}_0 \left(\sum_{i=1}^{T_0} f(X_i) \right) - \mathbb{E}_0 \left(\sum_{i=1}^{T_0} f(X_i^N) \right) \right| \leq 3\varepsilon.$$

Ce qui prouve (2.8).

En choisissant $f \equiv 1$, on obtient

$$\lim_{N \rightarrow +\infty} \pi_N(0) = \lim_{N \rightarrow +\infty} \frac{1}{\mathbb{E}(T_0^N)} = \frac{1}{\mathbb{E}(T_0)} = \pi(0),$$

puis pour tout x élément de S ,

$$\pi(x) = \frac{\mathbb{E} \left(\sum_{i=1}^{T_0} \mathbf{1}_x(X_i) \right)}{\mathbb{E}(T_0)} = \lim_{N \rightarrow +\infty} \frac{\mathbb{E} \left(\sum_{i=1}^{T_0^N} \mathbf{1}_x(X_i) \right)}{\mathbb{E}(T_0^N)} = \lim_{N \rightarrow +\infty} \pi_N(x).$$

□

Remarque 3. *La conclusion (2.7) reste vraie pour $\alpha = 1$. Par contre la convergence des mesures invariantes tombe en défaut. En effet, considérons pour tout naturel N supérieur à 1, une marche aléatoire (X_n^N) définie sur \mathbb{N} dont les probabilités de transition sont les suivantes :*

$$\forall n \neq 0 \quad P^N(n, n-1) = 1$$

$$P^N(0, N) = \frac{1}{N} \text{ et } P^N(0, 1) = 1 - \frac{1}{N}.$$

Et la marche déterministe (X_n) définie par

$$\forall n \neq 0 \quad P(n, n-1) = 1,$$

$$P(0, 1) = 1.$$

Nous avons convergence des probabilités de transition.

$$\mathbb{E}[T_0^N] = 2 \left(1 - \frac{1}{N} \right) + (N+1) \left(\frac{1}{N} \right) = 3 - \frac{1}{N}.$$

Les hypothèses de la proposition (2.2.4) sont vérifiées pour $\alpha = 1$. Or,

$$\pi_N(0) = \pi_N(1) = \frac{1}{3 - \frac{1}{N}}$$

$$\pi(0) = \pi(1) = \frac{1}{2}.$$

2.3 Limites fluides

Les limites fluides peuvent être utilisées pour analyser les performances globales d'un réseau de files d'attente dans des conditions de charge élevée. En considérant le système dans son ensemble comme un flux continu, les limites fluides permettent de simplifier la modélisation des réseaux de files d'attente complexes et d'obtenir des résultats théoriques sur les performances du système.

2.3.1 Renormalisation des processus

Définition 2.3. Si f est une fonction qui strictement positive sur S et $x \in S$, alors on note $(\| X \|_f (x, t))$ le processus défini par

$$\| X \|_f (x, t) = \frac{f(X(x, f(x)t))}{f(x)}, \quad (2.13)$$

pour $x \in S$ et $t \geq 0$. Si S est inclus dans un espace vectoriel sur \mathbb{R} , $(\bar{X}_f(x, t))$ désigne le processus $(X(x, t))$ renormalisé par la fonction f

$$\bar{X}_f(x, t) = \frac{1}{f(x)} X(x, f(x)t). \quad (2.14)$$

S'il n'y a pas d'ambiguïté sur le point initial x ou la fonction f , l'indice correspondant sera omis, i.e. les notions $(\bar{X}(x, t))$ et $(\| X \| (x, t))$ ou $(\bar{X}_f(t))$ et $(\| X \|_f (t))$ seront utilisées pour désigner ces processus.

Remarque 4. L'origine $\| X \|_f (x, 0) = 1$; si S est dans un espace vectoriel et si la fonction f est une norme sur l'espace S , alors pour $\| X \|_f (x, t) = f(\bar{X}_f(x, t))$ pour $t \geq 0$, le processus $(\bar{X}_f(x, t))$ part d'un état qui est de norme 1. La renormalisation consiste à accélérer le temps d'un facteur $f(X(0))$ tout en renormalisant en espace par $1/f(X(0))$.

Dans ce section on considère le cadre des processus de Markov à temps continu qui est similaire dans le cadre du temps discret : si (X_n) est une chaîne de Markov, le processus renormalisé est défini comme

$$\| X \|_f (x, t) = \frac{f(X_{\lfloor f(x)t \rfloor})}{f(x)},$$

si $X_0 = x \in S$ et $t \in \mathbb{R}_+$ (comme d'habitude, $\lfloor y \rfloor$ désigne la partie entière de $y \in \mathbb{R}$).

2.3.2 Limites fluides

Définition 2.4. Une limite fluide associée au processus de Markov $(X(t))$ et à une fonction positive f est un point d'accumulation des lois de probabilité des processus

$$\left\{ \left(\frac{f(X(x, f(x)t))}{f(x)} \right); x \in S \right\} = \{ (\| X \|_f (x, t)); x \in S \}$$

Sur l'espace des fonctions càdlàg muni de la topologie de Skorokhod.

Si l'espace d'état S est dans un espace vectoriel de dimension finie, le point d'accumulation des lois de probabilité de l'ensemble des processus

$$\left\{ \left(\frac{1}{f(x)} X(x, f(x)t) \right); x \in S \right\} = \{ (\bar{X}_f(x, t)); x \in S \}$$

est aussi appelé par abus de langage une limite fluide. Une limite fluide est donc une loi de probabilité Q d'un processus càdlàg. Quitte à agrandir l'espace de probabilité initial, une limite de fluide peut être représentée comme un processus $(W(t))$ de loi Q défini sur l'espace de probabilité de base. Les limites de fluides fournissent une expression asymptotique des trajectoires du processus de Markov, permettant ainsi de déterminer le comportement qualitatif de ce dernier pour de grandes valeurs initiales.

2.3.3 Relation entre renormalisation et ergodicité

On revient dans cette section au cas d'un processus de Markov général à valeurs dans un espace d'états dénombrable. Le résultat suivant est dû à Rybko et Stolyar [20] (quand le temps d'arrêt τ est déterministe).

Théorème 2.3.1. *S'il existe une fonction f positive sur S , des constants A, ε strictement positives et un temps d'arrêt τ tels que*

$$\limsup_{f(x) \rightarrow +\infty} \frac{\mathbb{E}_x(\tau)}{f(x)} \leq A, \quad (2.15)$$

$$\limsup_{f(x) \rightarrow +\infty} \mathbb{E}_x \left(\frac{f(X(\tau))}{f(x)} \right) \leq 1 - \varepsilon, \quad (2.16)$$

si la variable $f(X(1))$ est intégrable et si l'ensemble

$$F_K = \{x \in S / f(x) \leq K\}$$

est fini pour tout $K \geq 0$, donc le processus de Markov $(X(t))$ est ergodique.

Pour K assez grand, le temps d'atteinte T_{F_K} de F_K par $(X(t))$ satisfait l'inégalité

$$\frac{\mathbb{E}_x(T_{F_K})}{f(x)} \leq \frac{4A}{\varepsilon}, \quad (2.17)$$

Pour $x \in S$ tout tel que $f(x) > K$.

Démonstration. Il existe $K \geq 0$, tel que pour tout $x \in S$ vérifiant $f(x) \geq K$ les inégalités suivantes soient vérifiées,

$$\mathbb{E}_x(f(X(\tau))) \leq (1 - \varepsilon/2)f(x) \text{ et } \mathbb{E}_x(\tau) \leq 2Af(x).$$

Par conséquent si $f(x) > K$, on en déduit l'inégalité

$$\mathbb{E}_x(f(X(\tau))) - f(x) \leq -\frac{\varepsilon}{2}f(x) \leq -\frac{\varepsilon}{4A}\mathbb{E}_x(\tau).$$

Comme l'ensemble $\{x \in S / f(x) \leq K\}$ est fini, le théorème (3.2.2) dans Annexe montre que le processus $(X(t))$ est ergodique. La majoration de la moyenne de T_{F_K} est une conséquence de l'inégalité (3.29) de ce théorème. □

La fonction $t \rightarrow f(X(x, t))/f(x)$ part de 1 et la condition (2.16) impose qu'au bout d'un certain temps τ , elle soit en moyenne strictement plus petite que 1 si l'état initial est assez grand.

La proposition ci-dessous fait le lien entre les questions d'ergodicité et le comportement du processus renormalisé.

Proposition 2.3.1. *Si une fonction positive f sur S est telle que*

a) pour $K \geq 0$, l'ensemble $\{x \in S / f(x) \leq K\}$ est fini,

b) la variable $f(X(1))$ est intégrable, et s'il existe $T > 0$ et $\varepsilon > 0$ tels que

$$\limsup_{f(x) \rightarrow +\infty} \mathbb{E}_x(\|X\|_f(x, t)) = \limsup_{f(x) \rightarrow +\infty} \frac{\mathbb{E}_x(f(X(f(x)T)))}{f(x)} \leq 1 - \varepsilon,$$

le processus de Markov $(X(t))$ est ergodique.

La démonstration est dans le livre de Philippe Robert [19].

Le corollaire suivant montre si toute les limites fluides tendent vers 0 convenablement alors le processus de Markov est ergodique.

Corollaire 2.3.1. *Si les conditions a) et b) de la proposition précédente sont satisfaites et s'il existe T tel que*

c) *la variable $\|X\|_f(x, T)$ converge en loi vers 0 quand $f(x)$ tend vers l'infini,*

d) *la suite $(\|X\|_f(x, T); x \in S)$ est uniformément intégrable,*

le processus de Markov est ergodique. La condition c) peut être remplacée par

c') *le processus $(\|X\|_f(x, t))$ converge pour la topologie de Skorokhod vers $(x(t))$ quand $f(x)$ tend vers l'infini et $x(t) = 0$ pour tout $t \geq T/2$.*

La démonstration est dans [19].

2.3.4 Approximation fluide et stabilité

Dans cette partie, nous présentons une approche qui a été étudiée par Rybko et Stolyar [20] et Dai [5]. L'idée est basée sur l'étude des versions renormalisées (scaled) du processus de Markov. A cette fin, soit $(r_n, x_n)_{n \in \mathbb{N}}$ une suite de couples, où $r_n \in \mathbb{R}_+$ et $x_n \in \chi$ est une suite des états initiaux. On suppose que la suite de couples satisfait les conditions suivantes

$$\lim_{n \rightarrow \infty} r_n = \infty, \quad \limsup_{n \rightarrow \infty} \frac{\|q_n\|}{r_n} < \infty, \quad \lim_{n \rightarrow \infty} \frac{\|u_n\|}{r_n} = \lim_{n \rightarrow \infty} \frac{\|\nu_n\|}{r_n} = 0, \quad (2.18)$$

où q_n , u_n et ν_n désignent la longueur de la file d'attente, le temps d'inter-arrivées résiduel, et le temps de service résiduel, respectivement.

Dans la suite, nous considérons la famille $\dot{X} = \{X_n(t), t \geq 0, n \in \mathbb{N}\}$ des processus de Markov définie par

$$X_n(t) = \frac{1}{r_n} X^{x_n}(r_n t),$$

où l'exposant x_n exprime la dépendance à l'état initial $x_n \in \chi$.

Afin d'étudier la famille \dot{X} nous commençons à mettre l'accent sur les cumulatifs primitifs (E, S, R) .

Dans le lemme suivant nous rappelons les résultats de convergence des versions renormalisées des cumulatifs primitifs.

Lemme 2.3.1. *On suppose que la suite de couples $(r_n, x_n)_{n \in \mathbb{N}}$ satisfait (2.18). Alors, presque sûrement, quand $n \rightarrow \infty$,*

$$\frac{1}{r_n} E_k^{x_n}(r_n t) \rightarrow \alpha_k t \text{ c.u.c}^1, \quad \frac{1}{r_n} S_k^{x_n}(r_n t) \rightarrow \mu_k t \text{ c.u.c}, \quad \frac{1}{r_n} R^k([r_n t]) \rightarrow P_k t \text{ c.u.c}, \quad (2.19)$$

1. La suite $(x_n)_{n \in \mathbb{N}}$ dans $D(\mathbb{R}_+, \mathbb{R}^N)$ converge uniformément sur des ensembles compacts (c.u.c) vers $x \in D(\mathbb{R}_+, \mathbb{R}^N)$ si pour chaque $T > 0$, on a

$$\lim_{n \rightarrow \infty} \sup_{t \in [0, T]} \|x_n(t) - x_n(0)\| = 0.$$

où $[a]$ désigne la partie entière de $a \in \mathbb{R}$.

Théorème 2.3.2. *Pour chaque réseau de files d'attente HL, $(r_n, x_n)_{n \in \mathbb{N}}$ satisfaisant (2.18) et $\omega \in G$ il existe une sous-suite de couples $(r_{n_i}, x_{n_i})_{i \in \mathbb{N}}$ de telle sorte que*

$$\lim_{i \rightarrow \infty} \frac{1}{r_{n_i}} X^{x_{n_i}}(r_{n_i} t, \omega) = \bar{X}(t, \omega) \quad c.u.c. \quad (2.20)$$

Toute limite : $\bar{X}(\cdot) = (\bar{A}(\cdot), \bar{D}(\cdot), \bar{T}(\cdot), \bar{W}(\cdot), \bar{I}(\cdot), \bar{Q}(\cdot))$ obtenue à partir de l'équation (2.20) est appelée une limite fluide sous une certaine discipline de service.

L'ensemble de toutes les limites fluides associées à la trajectoire ω est désignée par $\mathfrak{FL}(\omega)$.

Donc, chaque fois qu'une limite fluide est considérée, nous avons $\bar{X}(\cdot) \in \mathfrak{FL}(\omega)$ pour certaine $\omega = \{a(n), s(n), \phi(n), \quad n \in \mathbb{N}\}$.

Un modèle de limite fluide selon Schönlein est défini comme suit :

Définition 2.5. *L'ensemble de toutes les limites fluides pour toutes les trajectoires d'échantillon ω est appelé un modèle de limite fluide (fluid limit model; \mathfrak{FLM}), c'est à dire,*

$$\mathfrak{FLM} = \{X(\omega), \quad \omega \in G\}.$$

Où G est l'ensemble de tous les chemins d'échantillon satisfaisant (2.19).

Conditions de stabilité d'un réseau de file d'attente multiclasse avec priorité

Le réseau fluide correspondant est un outil important pour étudier la stabilité des réseaux de files d'attente, il s'agit d'une simulation continue d'un réseau de files d'attente.

Le théorème proposé par Rybko et Stolyar [20] et étendu par Dai [7] déclare qu'un réseau de files d'attente est stable si son modèle fluide correspondant est stable. Cela a inspiré des recherches sur la stabilité des réseaux fluides. Dai et Vande Vate [11, 9] ont décrit la région de stabilité globale pour les réseaux à deux stations, et dans [8] les auteurs ont caractérisé la méthode utilisée dans [1, 2, 11, 9] aux réseaux à trois stations.

Dans ce chapitre, nous utilisons l'approche de la fonction de Lyapunov linéaire par morceaux pour établir les conditions globales de stabilité d'un réseau de files d'attente multi-classes avec priorité composé de N unités $N \geq 3$. Même si l'intensité du trafic à chaque unité du réseau est inférieure à un, la détermination de la région de stabilité globale d'un tel système est très difficile dans de mauvaises disciplines. Chaque unité du réseau a une intensité de trafic inférieure à 1. En fait, nous démontrons que dans les conditions normales de trafic habituelles :

$$\rho = \left(\alpha \sum_{k:\sigma(k)=i} m_k < 1 \right), \text{ pour } i = 1, 2, \dots, N,$$

de la condition supplémentaire suivante :

$$\alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} \leq 1$$

est suffisante pour assurer la stabilité globale.

3.1 Description du modèle et notations

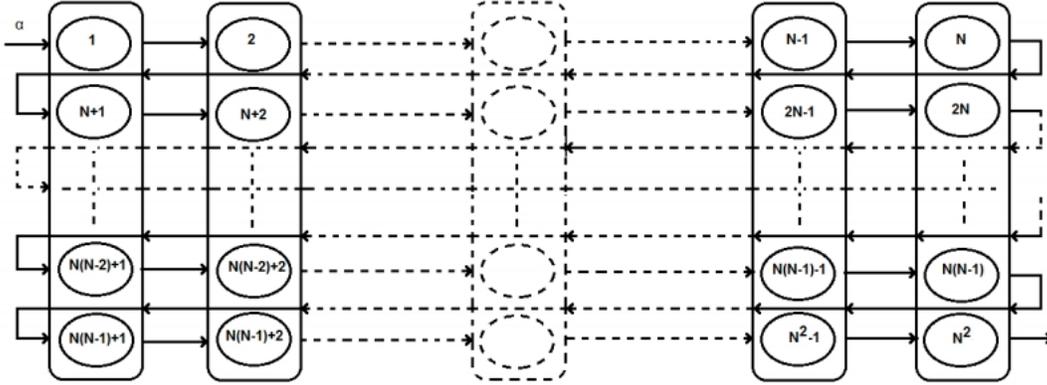


FIGURE 3.1 – Un réseau de file d'attente avec N unités et N^2 classes

Nous considérons un réseau de file d'attente avec priorité composé de N unités et N^2 classes. Un serveur peut servir N classes de clients dans chaque unité. Un fluide arrive de l'extérieur sur le réseau au taux α par unité de temps, ce qui nécessite un service de moyenne m_1 , puis il s'aligne sur la deuxième file d'attente demandant un service de moyenne m_2 , il est ensuite transféré dans la troisième unité où il sera servi une nouvelle fois, puis à la quatrième, jusqu'à la $N^{\text{ième}}$ unité, où il sera servi avec la moyenne m_N .

Après avoir été servi à la station N , le fluide revient à la station 1 et il est servi encore par chaque station alternativement avant qu'il quitte le système. Ainsi, chaque unité de fluide est traitée N^2 fois, N fois à chaque station, avant qu'il quitte le système.

Chaque fluide résidant dans la file $k, k = 1, 2, \dots, N^2$ est appelé fluide de classe k .

- Soit $Q_k(t)$ le niveau de fluide dans la classe k au le temps t .
- $T_k(t)$ est le temps de service accumulé $\sigma(k)$ consacré à la classe k dans l'intervalle $[0, t]$.
- On note : $U_i(t) = t - \sum_{k:\sigma(k)=i} T_k(t)$ est le temps passé cumulé à la station $i, i = 1, 2, \dots, N$ dans l'intervalle $[0, t]$.
- Les niveaux des classes $Q(\cdot) = (Q_k(\cdot))_{1 \leq k \leq N^2}$ et $T(\cdot) = (T_k(\cdot))_{1 \leq k \leq N^2}$ doivent satisfaire :

$$Q_k(t) = Q_k(0) + \mu_{k-1}T_{k-1}(t) - \mu_k T_k(t), \quad t \geq 0, \text{ pour } k = 1, \dots, N^2, \quad (3.1)$$

$$Q_k(t) \geq 0, \quad t \geq 0, \text{ pour } k = 1, \dots, N^2, \quad (3.2)$$

$$T_k(\cdot) \text{ est croissant, } k = 1, \dots, N^2, \quad (3.3)$$

$$U_i(.) \text{ est croissant, } k = 1, \dots, N^2, \quad (3.4)$$

- Avec, $\mu_k = 1/m_k$ est le taux de service pour la classe $k, k = 1, \dots, N^2$.
- $\mu_0 = \alpha$ est le taux des arrivées exogènes.
- $T_0(t) = t$ modélise le processus d'arrivée exogène.
- $\mu_k T_k(t)$ est la quantité de fluide à avoir quitté la classe k au temps t .

Toute solution $(Q(.), T(.))$ des équations (3.1)-(3.4) est une solution fluide. Une solution fluide $(Q(.), T(.))$ satisfaisante :

$$\int_0^{\infty} Z_i(t) dU_i(t) = 0, \quad i = 1, \dots, N, \quad (3.5)$$

où

$$Z_i(t) = \sum_{k:\sigma(k)=i} Q_k(t), \quad i = 1, \dots, N, \quad (3.6)$$

est dit non-idling.

Les équations (3.1)-(3.5) définissant un réseau fluide sous disciplines service de non-idling.

Pour toute solution fluide $(Q(.), T(.))$, $Q(.)$ est différentiable pour tout t dans $(0, \infty)$; t est un point régulier pour la solution fluide $(Q(.), T(.))$ si $T(.)$ est différentiable à l'instant t . Pour une fonction différentiable $h : [0, \infty] \rightarrow \mathbb{R}$, $\dot{h}(t)$ est la dérivée de h à l'instant t . Notons que (3.5) est équivalente à la condition $Z_i(t) > 0$ implique $\dot{U}_i(t) = 0$,

pour chaque point régulier t . Il a été montré dans [2] que chaque limite fluide est une solution fluide satisfaisant (3.1)-(3.5).

L'une des disciplines les plus courantes est celle de la discipline avec priorité. Ce qui signifie que le serveur ne peut pas travailler sur des classes de priorité inférieure à une unité que lorsque les besoins des classes de priorité supérieure sont satisfaits.

Chaque unité de notre réseau peut servir N classes. Soit π_i la classe de haute priorité à l'unité i sous la discipline de priorité statique π . Ainsi, la discipline de priorité spécifique considérée est la suivante :

- Si l'ordre de la station $i, (i = \overline{1, N})$ est impair, la classe k qui appartient à cette station a une priorité par rapport à la classe d'ordre $k - N$.
- Si l'ordre de la station $i, i = \overline{1, N}$ est pair, on a la classe k qui appartient à cette station, a une priorité par rapport à la classe d'ordre $k + N$.

Exemple 1.

$\pi_{\{13,9,5,1,2,6,10,14,15,11,7,3,4,8,12,16\}}$ désigne la discipline de service avec priorité qui donne :

- Dans la station 1 une priorité plus élevée à la classe 13 par rapport à la classe 9, et une priorité plus élevée à la classe 9 par rapport à la classe 5 et une priorité plus élevée à la classe 5 par rapport à la classe 1.
- Dans la station 2, donne une priorité supérieure à la classe 2 par rapport à la classe 6, une priorité plus élevée à la classe 6 par rapport à la classe 10 et une priorité plus élevée à la classe 10 par rapport à 14.
- Et dans la station 3 une priorité supérieure à la classe 15 par rapport à la classe 11, une priorité plus élevée à la classe 11 par rapport à la classe 7 et une priorité plus élevée à la classe 7 par rapport à 3.
- Et finalement dans la station 4, elle donne une une priorité plus élevée à la classe 4 par rapport à la classe 8 et une priorité plus élevée à la classe 8 par rapport à 12 et une priorité plus élevée à la classe 12 par rapport à 16

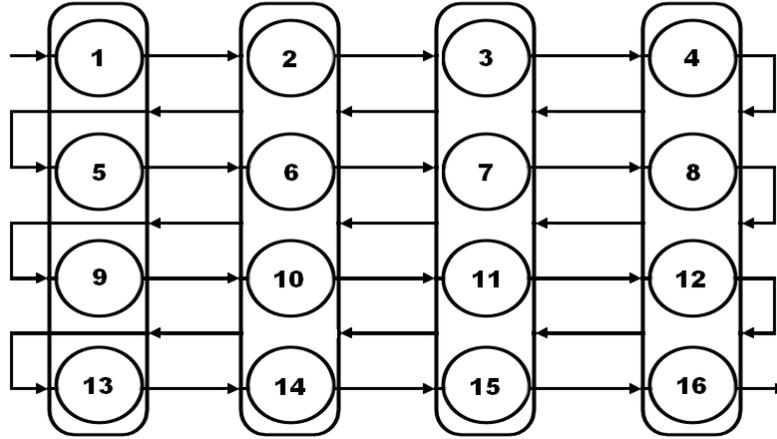


FIGURE 3.2 – Réseau de file dattente fluide composé de 4 stations et 16 classes de clients sous des disciplines de service avec priorité

Avec cette notation, notre réseau fluide sous la discipline de service avec priorité π nécessite les équations additionnelles suivantes :

$$\dot{T}_{\pi(i)}(t) = 1 \text{ Si } Q_{\pi(i)}(t) > 0, \quad i = 1, \dots, N, \quad (3.7)$$

pour chaque point régulier t de $T(\cdot)$. Chaque solution $(Q(\cdot), T(\cdot))$ de (3.1)-(3.5) et (3.7) est une solution fluide sous la discipline π .

Définition 3.1. *Le réseau de fluides est globalement stable s'il existe un temps $\sigma > 0$ tel que pour chaque solution fluide non idling $(Q(\cdot), T(\cdot))$ satisfaisant (3.1)-(3.5) avec $|Q(0)| = 1, Q(t) = 0$ pour tout $t \geq \sigma$.*

Définition 3.2. *Le réseau fluide sous une discipline de priorité π est stable s'il existe un temps $\sigma > 0$ tel que pour chaque solution fluide $(Q(\cdot), T(\cdot))$ satisfaisant (3.1)-(3.5) et (3.7) avec $|Q(0)| = 1, Q(t) = 0$ pour tout $t \geq \sigma$.*

Définition 3.3. Une solution fluide $(Q(\cdot), T(\cdot))$ est instable s'il n'existe pas de $\sigma > 0$ tel que $Q(t) = 0$ pour tout $t \geq \sigma$.

Définition 3.4. Pour un $\alpha > 0$ donné, la région de stabilité globale du réseau fluide est l'ensemble de temps de service positifs $m = (m_k)$ pour lesquels le réseau de fluides est globalement stable.

Définition 3.5. Pour un $\alpha > 0$ donné et une discipline de priorité π , la région de stabilité du réseau fluide sous la discipline est l'ensemble des temps de services positifs $m = (m_k)$ pour lesquels le réseau fluide sous la discipline est stable.

Nous disons que la condition habituelle est satisfaite si l'intensité du trafic pour chaque unité de temps est inférieure à 1, c'est-à-dire :

$$\rho = \alpha \sum_{k:\sigma(k)=i} m_k < 1 \text{ pour chaque } i = 1, \dots, N. \quad (3.8)$$

Nous avons maintenant le système de contraintes linéaires suivant, qui est lié à une fonction de Lyapunov linéaire par morceaux pour notre réseau de fluides à N unités :

$$\alpha \sum_{k:\sigma(k)=i} x_k < x_k \mu_k, \quad i = 1, \dots, N. \quad (3.9)$$

$$\left(\sum_{k:\sigma(k)=1} x_k \right) - x_1 \leq \sum_{k:\sigma(k)=N} x_k \quad (3.10)$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i \leq \sum_{k:\sigma(k)=i-1} x_k - x_i, \quad i = 2, \dots, N, \quad (3.11)$$

$$\sum_{k:\sigma(k)=i} x_k \leq \sum_{k:\sigma(k)=i-1} x_k \quad i = 2, \dots, N. \quad (3.12)$$

Le système de contraintes linéaires (3.9)-(3.12) dérivé de notre fonction de Lyapunov linéaire par morceaux fournit des conditions suffisantes pour assurer la stabilité globale du système.

3.2 Résultat principal

Dans cette section, nous caractérisons la région de stabilité globale à l'aide des fonctions de Lyapunov linéaires par morceaux (voir [3],[21]). Étant donné $x = (x_k) > 0$ et une solution fluide $Q(\cdot)$, soit :

$$h_i(x, Q(t)) = \sum_{k:\sigma(k)=i} x_k Q_k^+(t), \quad i = 1, \dots, N,$$

avec $Q_k^+(t) = \sum_{l=1}^k Q_l(t)$, et $h(x, Q(t)) = \max \{h_i(x, Q(t))\}$, $i = 1, \dots, N$.

Donc $h(Q(t))$ est une fonction linéaire par morceaux de $Q(t) = (Q_k(t))$.

La fonction de Lyapunov linéaire par morceaux h est dite fonction de Lyapunov pour la stabilité globale du modèle fluide s'il existe $\varepsilon > 0$ tel que pour chaque solution de fluide stable $(Q(\cdot), T(\cdot))$ satisfaisant (3.1)-(3.5),

$$\frac{dhQ((t))}{dt} \leq -\varepsilon \quad (3.13)$$

$\forall t > 0$ qui est régulier pour $T(\cdot)$ et $h(Q(t))$ avec $|Q(t)| > 0$.

Soit $m > 0$ un vecteur de temps de service pour lequel il existe une fonction de Lyapunov linéaire par morceaux h satisfaisant (3.13).

Il découle du lemme (2.2) de Dai et Weiss [10] que $h(Q(t)) = 0$ pour tout $t \geq h(Q(0))/\varepsilon$, ou $Q(t) = 0$ pour tout $t \geq h(Q(0))/\varepsilon$.

Le lemme suivant suggère une manière de construire des fonctions de Lyapunov linéaires par morceaux, il a été introduit par Botvich et Zamatina [2] pour un réseau à deux stations.

Il a été généralisé indépendamment par Dai et Weiss [10], Down et Meyn [11] et [8].

En utilisant ce résultat, nous obtenons

Lemme 3.2.1. *Supposons qu'il existe $x = (x_k) > 0$, $t_0 \geq 0$ et $\varepsilon > 0$ de sorte que pour chaque solution fluide non instable $(Q(\cdot), T(\cdot))$ et chaque point régulier $t > t_0$ de $T(\cdot)$, les conditions suivantes s'appliquent pour chaque $i = 1, 2, \dots, N$.*

$$\frac{dh_i(x, Q(t))}{dt} \leq -\varepsilon \text{ pour chaque } Z_i(t) > 0, \quad (3.14)$$

$$h_i(x, Q(t)) \leq \max \{h_j(x, Q(t)) : j \in \{1, \dots, N\}, j \neq i\} \text{ pour chaque } Z_i(t) = 0, \quad (3.15)$$

$$\max \{h_j(Q(t)) : j \in \{1, \dots, N\}, j \neq i\} \leq h_i(Q(t)) \text{ pour chaque } \sum_{j \neq i} Z_j(t) = 0. \quad (3.16)$$

Alors h est une fonction de Lyapunov linéaire par morceaux.

Proposition 3.2.1. *S'il existe $x = (x_k) > 0$ satisfaisant les contraintes linéaires (3.9)-(3.12), alors il existe $\varepsilon > 0$ tel que (3.14)-(3.16) sont valables et donc, h est une fonction de Lyapunov linéaire par morceaux.*

Démonstration. Soit $t_0 = 0$ et $x = (x_k) > 0$ satisfaisant les contraintes linéaires (3.9)-(3.12), on définit $\varepsilon > 0$ comme étant soit le minimum des termes suivants

$$x_k \mu_k - \alpha \left(\sum_{k:\sigma(k)=i} x_k \right), \quad i = 1, \dots, N.$$

On remarque que le niveau de fluide dans la classe k avec $k = 1, \dots, N^2$ est :

$$Q_k^+ = Q_k^+(0) + \alpha t - \mu_k T_k(t).$$

Ainsi

$$h_1(x, Q(t)) = h_1(0) + \alpha t \left(\sum_{k:\sigma(k)=1} x_k \right) - \sum_{k:\sigma(k)=1} x_k \mu_k T_k(t)$$

et

$$\frac{dh_1(Q(t))}{dt} = \alpha \left(\sum_{k:\sigma(k)=1} x_k \right) - \sum_{k:\sigma(k)=1} x_k \mu_k \dot{T}_k(t)$$

Nous établissons (3.14).

Si $Z_1 > 0$, cela découle de (3.7), $\sum_{k:\sigma(k)=1} \dot{T}_k(t) = 1$ ainsi $\dot{h}_1(t) < -\varepsilon$.

Nous suivons une analyse similaire pour $i = (\overline{2}, \overline{N})$.

Ensuite, nous établissons (3.15).

- Lorsque $Z_1(t) = 0$, l'équation (3.10) implique que $h_1(Q(t)) < h_N(Q(t))$.
- Lorsque $Z_i(t) = 0, i = (\overline{2}, \overline{N})$, l'équation (3.11) implique que $h_i(Q(t)) < h_{i-1}(Q(t))$.

Enfin, nous établissons (3.16).

Lorsque $\sum_{j \neq i} Z_j(t) = 0, j = (\overline{1}, \overline{N})$, les équations (3.11)-(3.12) vérifient que $h_j(Q(t)) < h_i(Q(t))$.

Le résultat suivant établit des conditions suffisantes pour assurer la stabilité globale de notre réseau. □

Proposition 3.2.2. : Si

$$\alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} \leq 1, \quad (3.17)$$

$$\alpha \sum_{k:\sigma(k)=i} m_k < 1, \quad i = 2, \dots, N, \quad (3.18)$$

le réseau fluide est globalement stable.

Démonstration. Soit $(Q(\cdot), T(\cdot))$ une solution fluide non ralentie avec $|Q(0)| = 1$. Soit

$$f_1(t) = \sum_{k:\sigma(k)=1} \sum_{j=1}^k m_k Q_j(t)$$

est la charge de travail totale de la première unité à l'instant t . D'après (3.1), nous obtenons :

$$f_1(t) = f_1(0) + \alpha t \left(\sum_{k:\sigma(k)=1} m_k \right) - \sum_{k:\sigma(k)=1} T_k.$$

Pour tout t régulier avec $Z_1(t) > 0$, par (3.5), $\dot{f}_1(t) = \left(\alpha \sum_{k:\sigma(k)=1} m_k \right) - 1$.

Comme $\alpha \sum_{k:\sigma(k)=1} m_k < 1$, il existe un t_0 positif avec :

$$t_0 \leq \frac{f_1(0)}{1 - \alpha \sum_{k:\sigma(k)=1} m_k} \leq \frac{\sum_{k:\sigma(k)=1} k m_k}{1 - \alpha \sum_{k:\sigma(k)=1} m_k}$$

tel que $Z_1(t_0) = 0$. Supposons que (3.17) soit valable. Nous montrons ensuite que $Z_1(t) = 0$ pour $t \geq t_0$. Pour ce faire, posons :

$$f_2(t) = \sum_{k:\sigma(k)=1} m_k Q_k(t)$$

est la charge de travail immédiate de l'unité 1. Alors, à partir de (3.1)-(3.4),

$$f_2(t) = f_2(0) + \alpha m_1 t - \sum_{k:\sigma(k)=1} T_k(t) + \sum_{k:\sigma(k)=1} m_k \mu_{k-1} T_{k-1}(t)$$

et, pour tout t régulier avec $f_2(t) > 0$,

$$\dot{f}_2(t) = \alpha m_1 + \sum_{k:\sigma(k)=1} \frac{m_k}{m_{k-1}} - 1 \leq 0.$$

Ainsi, f_2 n'est pas croissante. Puisque $\dot{f}_2(t_0) = 0$, on a $Z_1(t) = 0$ pour $t \geq t_0$.

Il reste maintenant à montrer que $\sum_{i=2}^N Z_i(t) = 0$ pour chaque temps $t \geq t_1 \geq t_0$ et que le réseau est globalement stable.

Si nous considérons des temps $t \geq t_0$ et nous spécialisons le lemme (3.2.1) au cas où $Z_1(t) = 0$ et $\dot{Q}_1(t) = \dots = \dot{Q}_{N(N-1)+1}(t) = 0$.

Tout d'abord, observons que puisque $Z_1(t) = 0$ pour $t \geq t_0$, (3.14)-(3.16) sont satisfaites pour $i = 1$.

Ensuite, en rappelant que (3.16) implique (3.15) dans notre réseau, on voit qu'il nous reste les conditions :

$$\frac{dh_i(x, Q(t))}{dt} \leq -\varepsilon \text{ lorsque } Z_i(t) > 0, \quad i = 2, \dots, N \quad (3.19)$$

$$h_k(Q(t)) = \max \{h_i(Q(t)), h_k(Q(t)), i \in \{1, \dots, N\} \setminus \{k, j\}, \quad k = 2, \dots, N\}$$

$$\leq h_j(Q(t)), j \in \{1, \dots, N\} \setminus \{k, i\} \text{ lorsque } \sum_{l=1}^{N \setminus j} Z_l = 0 \quad (3.20)$$

Ensuite, par le lemme (3.2.1), nous pouvons facilement montrer que (3.19)-(3.20) et donc (3.14)-(3.16) sont valables s'il existe $(x_2, x_3, \dots, x_{N^2-1}, x_{N^2}) > 0$ satisfaisant

$$\alpha \sum_{k:\sigma(k)=i} x_k < \mu_k x_k, \quad i = 2, \dots, N \quad (3.21)$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i < \sum_{k:\sigma(k)=N} x_k, \quad i = 2, \dots, N-1 \quad (3.22)$$

$$\sum_{k:\sigma(k)=i} x_k < \sum_{k:\sigma(k)=i-1} x_k, \quad i = 2, \dots, N \quad (3.23)$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i < \sum_{k:\sigma(k)=i-1} x_k, \quad i = 3, \dots, N \quad (3.24)$$

$$\left(\sum_{k:\sigma(k)=i} x_k \right) - x_i < \sum_{k:\sigma(k)=i-1} x_k - x_i, \quad i = 3, \dots, N \quad (3.25)$$

Alors, étant donné $(x_1, x_2, \dots, x_{N^2}) > 0$, laissez :

$$y_i^k = \frac{x_k}{\sum_{k:\sigma(k)=i} x_k}, \quad i = 1, \dots, N$$

Et $(x_1, x_2, \dots, x_{N^2}) > 0$ satisfait (3.9)-(3.12) si $(y_1, \dots, y_N, x_{N+1}, \dots, x_{N^2}) > 0$ satisfait

$$\alpha m_k < y_i^k, \quad k = 1, \dots, N^2, \quad i = 1, \dots, N. \quad (3.26)$$

Il existe donc $x > 0$ satisfaisant (3.21)-(3.25) si et seulement si les conditions de circulation habituelles (3.18) aux unités $2, 3, \dots, N$ sont satisfaites.

Par conséquent, la proposition découle du lemme (3.2.1). □

Conclusion

L'étude des réseaux de files d'attente fluides multi-classes est un domaine de recherche actif qui fournit des outils mathématiques puissants pour modéliser, analyser et optimiser les performances des réseaux de télécommunications modernes.

Dans ce travail, on a étudié les résultats obtenus dans l'article de Dai et Vande Vate [6] qui a prouvé que les fonctions de Lyapunov linéaires par morceaux utilisées dans ce mémoire caractérisent la région de stabilité globale de réseaux de files d'attente multi-classes de N stations et N^2 classes de clients. Ainsi, on a introduit le concept de stabilité pour le modèle fluide associé au réseaux correspondant.

Nous souhaitons que ce mémoire soit une bonne initiation pour nous à la poursuite de nos études et nos recherches dans le domaine des probabilités et de la statistique.

Annexe

Définition 3.6. (*Endomorphisme*). Un endomorphisme T de (Ω, \mathcal{F}, P) est une application de Ω dans Ω telle que

1. T est une application mesurable ;
2. T préserve la probabilité P : pour toute fonction f mesurable positive sur l'ensemble Ω ,

$$\int_{\Omega} f(T(\omega))dP(\omega) = \int_{\Omega} f(\omega)dP(\omega)$$

Définition 3.7. (*Endomorphismes induits*). Si $A \in \mathcal{F}$ tel que $P(A) > 0$, l'opérateur T_A de A dans A est appelé opérateur induit ,

$$T_A : \omega \rightarrow T^{v_A(\omega)}(\omega).$$

Si T est un endomorphisme de (Ω, \mathcal{F}, P) , l'opérateur T_A est un endomorphisme de l'espace de probabilité (A, \mathcal{F}_A, P_A) .

Définition 3.8. [19] Une fonction f d'un intervalle J dans \mathbb{R} est de Lipschitz d'ordre $\alpha > 0$ s'il existe une constante C telle que pour tout $x, y \in J$

$$|f(x) - f(y)| \leq C|x - y|^\alpha.$$

Proposition 3.2.3. pour $T > 0$ il existe une constante K_T telle que, si Y et Y' sont des fonctions càdlàg telles que $Y(0) \geq 0$ et $Y'(0) \geq 0$, alors

$$\| X_Y - X_{Y'} \|_{\infty, T} \leq \| Y - Y' \|_{\infty, T}, \quad (3.27)$$

et

$$\| R_Y - R_{Y'} \|_{\infty, T} \leq \| Y - Y' \|_{\infty, T}, \quad (3.28)$$

en notant $\| Z \|_{\infty, T} = \sup_{0 \leq s \leq T} \max_{1 \leq i \leq d} |Z_i(s)|$ si $Z(t) = (Z_i(t))$.

Si les coordonnées d'une fonction càdlàg Y sont de Lipschitz d'ordre α sur l'intervalle $[0, T]$, il en va de même pour R_Y et X_Y .

Théorème 3.2.1. (*Théorème de représentation de Skorokhod*).

Si une suite de probabilités (\mathbb{P}_n) sur $D([0, T], \mathbb{R}^d)$ converge vers la probabilité \mathbb{P} , il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{Q})$ sur lequel sont définis des processus càdlàg $(Y_n(t)), n \geq 1$ et $(Y(t))$ tels que, pour $n \geq 1$, la loi de $(Y_n(t))$ soit \mathbb{P}_n , P la loi de $(Y(t))$ et \mathbb{Q} -presque sûrement $(Y_n(t))$ converge, pour la topologie de Skorokhod, vers $(Y(t))$ quand n tend vers l'infini.

Théorème 3.2.2. *S'il existe une fonction $f : S \rightarrow \mathbb{R}_+$ des constantes $K, \gamma > 0$ et un temps d'arrêt τ intégrable tels que si $f(x) > K$,*

$$\mathbb{E}_x(f(x(\tau))) - f(x) \leq \gamma \mathbb{E}_x(\tau),$$

en notant $F = \{x/f(x) \leq K\}$, le temps d'atteinte T_F de F est intégrable et

$$\mathbb{E}_x(T_F) \leq f(x)/\gamma \tag{3.29}$$

pour $x \notin F$. Si de plus F est fini et $\mathbb{E}_x(f(X(1))) < +\infty$ pour tout $x \in S$, le processus de Markou $(X(t))$ est ergodique

Espace d'état

nous nous limiterons aux réseaux de files d'attente HL(head-of-the-line)(au sein de chaque classe, les clients sont servis dans l'ordre (first in first out) (FIFO)). L'évolution du réseau de files d'attente multi-classes HL sera décrite par un processus stochastique $\{X = X(t), t \geq 0\}$ et son espace d'état correspondant désigné par $(\chi, B(\chi))$ où

$$\chi = \{x : x = (y, z) = (((k, w), u, v), z) \in (\mathbb{Z} \times \mathbb{R})^\infty \times \mathbb{R}^{|\varepsilon|} \times \mathbb{R}^K \times \mathbb{R}^K\}, \tag{3.30}$$

et χ est muni d'une métrique d . Alors $(\mathbb{Z} \times \mathbb{R})^\infty$ est l'ensemble des variables finies à valeurs dans $(\mathbb{Z} \times \mathbb{R})$, et $B(\chi)$ est la norme de Borel σ -algèbre de χ induite par la métrique d . Avant d'introduire la métrique, nous considérons l'espace d'état χ . La signification de chacune de ses composantes est définie par :

- $(k, w) \in (\mathbb{Z} \times \mathbb{R})^\infty$: L'ordre global des clients du réseau est donnée par les couples de variables $(k, w) = ((k_1, w_1), (k_2, w_2), \dots, (k_l, w_l))$. La première entrée $k_i \in \{1, 2, \dots, K\}$ désigne la classe actuelle du $i^{\text{ème}}$ client et la seconde entrée $w_i \geq 0$ reflète le temps écoulé depuis que le client i est rentré dans la classe k_i . Notons que dans chaque classe le plus ancien client (le premier dans le couple (k, w)) c'est le client qui a la plus grande deuxième coordonnée. Dans le cas de deux ou plusieurs clients avec la deuxième coordonnée identique, l'ordre des clients est croissant par rapport à leur classe. Le nombre de clients de chaque classe est désigné par

$$q = (q_1, \dots, q_K),$$

et

$$\|q\|_1 = \sum_{k=1}^K q_k,$$

est le nombre total de clients dans l'espace d'état.

- $u \in \mathbb{R}^{|\varepsilon|}$: cette composante représente le temps d'intérrarrivées résiduel, c'est à dire, la coordonnée $u_k > 0$ est le temps restant avant la prochaine arrivée d'un client de classe $k \in \varepsilon$ de l'extérieur du réseau.

- $\nu \in \mathbb{R}^K$: cette composante représente le temps de service résiduel, c'est à dire, la coordonnée ν_k désigne le temps de service restant pour le client le plus ancien de la classe k , où $\nu_k \geq 0$ et

$\nu_k = 0$ seulement si $q_k = 0$.

• $z \in [0, 1]^k$: la composante z_k désigne la proportion de service de la station $c(k)$ que reçoit la plus ancienne classe k de client, pendant que d'autres clients de la classe k ne reçoivent pas de service. Cela représente la propriété HL. Pour chaque station j , nous avons : Si $\sum_{k \in c(j)} q_k > 0$, alors

$\sum_{k \in c(j)} z_k = 1$, où $z_k = 0$ si $q_k = 0$. Si la station j est vide, c'est à dire $\sum_{k \in c(j)} q_k = 0$ alors $\sum_{k \in c(j)} z_k = 0$.

Maintenant, nous introduisons une métrique d sur χ qui résume la différence de chaque composante. C'est à dire, pour $x, \hat{x} \in \chi$. Nous définissons

$$d(x, \hat{x}) = \sum_{i=1}^{\infty} \min\{|k_i - \hat{k}_i| + |w_i - \hat{w}_i| + |z_i - \hat{z}_i|, 1\} + \sum_{k \in \varepsilon} |u_k - \hat{u}_k| + \sum_{k=1}^K |\nu_k - \hat{\nu}_k|.$$

L'espace métrique (d, χ) possède les propriétés suivantes.

Proposition 3.2.4. [4] *L'espace métrique (χ, d) est séparable et localement compact. En particulier, χ est un espace Lusinien.*

Remarque 5. *L'espace métrique (χ, d) n'est pas complet. Une métrique complète peut être obtenue en ajoutant un terme approprié dans chacun des deuxième et troisième somme sur d , voir (Bramson [4] Section 4.1).*

Alors, l'espace d'état est mesurable. Avec un léger abus de notation, nous appelons

$$|x| = \|q\| + \|u\| + \|\nu\|,$$

la norme sur χ . En outre, χ est muni de la topologie naturelle induite et $\{x \in \chi : |x| \leq \kappa\}$ est un compact de χ pour chaque $\kappa > 0$

Bibliographie

- [1] D. Bertsimas, D. Gamarnik, and J. N. Tsitsiklis. Stability conditions for multiclass fluid queueing networks. *IEEE Transactions on Automatic Control*, 41(11) :1618–1631, 1996.
- [2] D. Botvich and A. Zamyatin. Ergodicity of conservative communication networks. rapport de recherche 1772. *INTRA*, October, 1992.
- [3] A. A. BOUCHENTOUF et al. Stability condition of a priority queueing system. *Mathematical Sciences and Applications E-Notes*, 1(2) :165–172, 2013.
- [4] M. Bramson. *Stability of queueing networks*. Springer, 2008.
- [5] H. Chen. Fluid approximations and stability of multiclass queueing networks : work-conserving disciplines. *The Annals of Applied Probability*, pages 637–665, 1995.
- [6] J. Dai and J. H. Vande Vate. Global stability of two-station queueing networks. In *Stochastic Networks*, pages 1–26. Springer, 1996.
- [7] J. G. Dai. On positive harris recurrence of multiclass queueing networks : a unified approach via fluid limit models. *The Annals of Applied Probability*, 5(1) :49–77, 1995.
- [8] J. G. Dai, J. J. Hasenbein, and J. Vande Vate. Stability of a three-station fluid network. *Queueing Systems*, 33 :293–325, 1999.
- [9] J. G. Dai and J. Vande Vate. The stability of two-station multitype fluid networks. *Operations Research*, 48(5) :721–744, 2000.
- [10] J. G. Dai and G. Weiss. Stability and instability of fluid models for reentrant lines. *Mathematics of Operations Research*, 21(1) :115–134, 1996.
- [11] D. Down and S. P. Meyn. Piecewise linear test functions for stability and instability of queueing networks. *Queueing Systems*, 27(3) :205–226, 1997.
- [12] M. Duflo. *Algorithmes stochastiques*, volume 23. Springer, 1996.
- [13] V. Dumas. A multiclass network with non-linear, non-convex, non-monotonic stability conditions. *Queueing systems*, 25 :1–43, 1997.
- [14] G. Fayolle, V. A. Malyshev, and M. V. Menshikov. *Topics in the constructive theory of countable Markov chains*. Cambridge university press, 1995.
- [15] Y. Filonov. A criterion for the ergodicity of discrete homogeneous markov chains. *Ukrain. Mat. Zh*, 41 :1421–1422, 1989.
- [16] R. Has. minskii. stochastic stability of differential equations, volume 7 of monographs and textbooks on mechanics of solids and fluids : Mechanics and analysis. *Sijthoff & Noordhoff, Alphen aan den Rijn*, 1980.

- [17] J. Lamperti. Criteria for the recurrence or transience of stochastic process. i. *Journal of Mathematical Analysis and applications*, 1(3-4) :314–330, 1960.
- [18] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [19] P. Robert. *Réseaux et files d’attente : méthodes probabilistes*, volume 35. Springer Science & Business Media, 2000.
- [20] A. N. Rybko and A. L. Stolyar. Ergodicity of stochastic processes describing the operation of open queueing networks. *Problemy Peredachi Informatsii*, 28(3) :3–26, 1992.
- [21] H. SAKHI et al. *Analyse dun réseau fluide multi-classe*. PhD thesis, 2017.
- [22] B. Sericola. *Markov chains : theory and applications*. John Wiley & Sons, 2013.
- [23] S. Smale and M. W. Hirsch. *Differential equations, dynamical systems, and linear algebra*, volume 60. Elsevier, 1974.

ملخص

سنقوم في هذه المذكرة بدراسة استقرار شبكة من الطوابير المائعة متعددة الفئات باستخدام دالة ليابونوف الخطية، نضع شرطا كافيا لضمان استقرار نظام الطوابير و الذي يتألف من N محطة و N^2 فئة من العملاء . تحديد منطقة الإستقرار الكلي في مثل هذا النظام صعب جدا في ظل خدمة معينة حتى لو كانت كثافة حركة المرور في كل وحدة من الشبكة أقل من 1

Résumé

Dans ce mémoire on va étudier la stabilité d'un réseaux de files d'attente fluides multi-classes. En utilisant la fonction de Lyapunov linéaire par morceaux, nous établissons une condition suffisante pour assurer la stabilité globale d'un système de files d'attente multi-classe composé de N stations et N^2 classes de clients sous des disciplines de services avec priorités. La détermination de la région de la stabilité globale d'un tel système est très difficile sous certaines disciplines de service, même si l'intensité du trafic à chaque unité du réseau est moins que 1.

Abstract

In this memory we will study the stability of a multi-class fluid queuing network. Using the linear Lyapunov function by morceaux, we establish a sufficient condition to ensure the global stability of a stability of a multi-class queuing system composed of N stations and N^2 customer classes under prioritized service disciplines. Determining the mination of the global stability region of such a system is very difficult under certain service disciplines, even if the traffic intensity at each network unit network unit is less than 1.