



HAL
open science

Propagation de la lumière dans les tissus biologiques par méthode de Monte Carlo par Chaînes de Markov

Tom Rohmer

► **To cite this version:**

Tom Rohmer. Propagation de la lumière dans les tissus biologiques par méthode de Monte Carlo par Chaînes de Markov. *Méthodologie [stat.ME]*. 2011. dumas-00623050

HAL Id: dumas-00623050

<https://dumas.ccsd.cnrs.fr/dumas-00623050v1>

Submitted on 14 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mémoire de Master 2 Statistique
Septembre 2011
Université de Strasbourg

Rohmer Tom

Propagation de la lumière dans les tissus biologiques par Méthode de Monte Carlo par Chaîne de Markov

Encadrants :

Samy Tindel	Professeur à l'IECN
Sandie Ferrigno	Maître de conférences à l'IECN

Effectué à l'Institut Elie Cartan de Nancy



Remerciements

Je remercie tout d'abord, Samy Tindel, professeur à l'Institut Elie Cartan de Nancy (IECN), et Sandie Ferrigno, maître de conférences à l'IECN, pour m'avoir permis de faire ce stage et m'avoir encadré tout au long de ces mois.

Je tiens également à remercier l'ensemble de l'équipe BIGS pour leurs aides précieuses, tout particulièrement Magalie Thomassin, maître de conférences au Centre de Recherche en Automatique de Nancy, pour son grand investissement dans la partie physique et programmation de mon sujet.

De même, je remercie Nicolas Poulain maître de conférences à l'université de Strasbourg, pour la relecture attentive et minutieuse de mon mémoire.

Enfin, un grand merci à Safaa El Sayed et Roger Nakad, doctorants de l'IECN, pour leur grand soutien tout au long de ce stage, et également à l'ensemble du personnel pour leur accueil chaleureux.

Notations

E	Espace des Etats
$\mathcal{B}(E)$ et $\mathcal{P}(E)$	L'ensemble des boréliens de E et l'ensemble des partitions de E
$ E $	Cardinal de E
\mathcal{C}^b	L'ensemble des fonctions continues bornées
μ	Loi de X_0 de la chaîne
$\mathbb{E}_\mu, \mathbb{P}_\mu$	Espérance et probabilité conditionnelle à la loi μ de X_0
$\mathbb{E}_X, \mathbb{P}_X$	Espérance et probabilité conditionnelle à la valeur prise par la variable aléatoire X
$\mathbf{1}(X_k = j)$	Fonction indicatrice, valant 1 si $X_k = j$ et 0 sinon.
$l^1(\pi)$	L'ensemble des fonctions f de $E \rightarrow \mathbb{R}$ telles que $\sum_{j \in E} f(j) \pi_j < +\infty$
f^+	$\max(0, f)$
f^-	$\max(0, -f)$
$\mathcal{L}(X)$	Loi de X
\mathcal{S}^2	Sphère unité.

Table des matières

Remerciements	iii
Notations	v
Introduction	1
1 Chaînes de Markov	3
1.1 Généralités sur les chaînes de Markov	3
1.2 Classification des états	8
1.3 Classification des chaînes	11
1.4 Temps d'arrêt et Propriétés de Markov	14
1.5 Invariance et réversibilité	16
1.6 Théorèmes ergodiques	18
2 Méthode de Monte-Carlo par Chaîne de Markov	25
2.1 Introduction	25
2.2 Algorithme de Metropolis indépendant	26
2.3 Algorithme de Metropolis généralisé	26
2.4 Exemples	29
2.4.1 Simulation d'une loi de Poisson	29
2.4.2 Modèle des sphères dures	30
2.5 Vitesse de convergence	32
2.5.1 Étude du spectre des matrices de transitions réversibles par rapport à π	33
2.5.2 Vitesse de convergence des noyaux réversibles	35
2.5.3 Formes de Dirichlet et trou spectral	37
3 Metropolis pour la synthèse d'image	41
3.1 Metropolis Light Transport	41
3.1.1 Equations de transport de la lumière	41
3.1.2 Estimation de l'intensité de lumière m_j	42
3.2 Influence du milieu	45
3.3 Algorithme de Metropolis	48
3.4 Mutations bidirectionnelles	50
4 Propagation de la lumière dans le cerveau	53
4.1 Mise en place de l'algorithme	53
4.2 Algorithme	55
4.3 Conclusion	56

Conclusion	59
Bibliographie	61
A Codes R du chapitre 2	63
A.1 loi de Poisson	63
A.2 Modèle des sphères dures	63
B codes matlab du chapitre 4	65
B.1 Génération du premier sommet :	65
B.2 Génération des sommets suivants :	65
B.3 Densité de probabilité d'obtenir un nouveau sommet	66
B.4 Choix du nombre de pont l_d à supprimer	66
B.5 Probabilité d'obtenir un l_d donné	66
B.6 Choix du nombre de pont l_a à ajouter, en fonction de l_d	67
B.7 Probabilité d'obtenir un l_a donné conditionnellement à l_d	67
B.8 Echantillonner un nouveau chemin dans la loi Q , sachant un chemin X_0	68
B.9 calcul de la probabilité d'obtenir ce nouveau chemin	69
B.10 calcul de f	70
B.11 Algorithme de Metropolis	70

Introduction

L'équipe BIGS (*Biology, Genitics and Statistics*) dans laquelle j'ai été accueilli au sein de l'institut Elie Cartan de Nancy, travaille en autres sur la mise en place de traitement contre le cancer à l'aide de la thérapie photodynamique. Il s'agit d'un traitement résidant sur trois facteurs : un agent photodynamique (médicament photoactivable), de la lumière et de l'oxygène.

Mon objectif à travers ce mémoire sera une étude approfondie des chaînes de Markov et des algorithmes de Monte Carlo par Chaînes de Markov afin d'appliquer ces connaissances pour simuler la propagation de la lumière dans un tissu du cerveau supposé homogène.

En effet l'objectif de l'équipe BIGS est de pouvoir simuler les effets de la lumière (son comportement dynamique et son intensité) dans les tissus d'un cerveau en présence d'une tumeur, afin d'aider les médecins et les biologistes à déterminer le traitement adéquat (type de diffuseur de lumière, placement de la fibre optique, dose de lumière). Pour optimiser le traitement, il faut s'assurer que l'ensemble de la tumeur sera irradiée par la lumière selon une dose adéquate; une trop forte dose de lumière provoquerait des effets nocifs et une trop faible pourrait ne pas détruire la tumeur dans son intégralité.

La propagation de la lumière à travers un milieu peut être modélisée par l'équation de transfert radiatif (ERT) issue de la physique, que l'on pourra retrouver dans [PAULY 00]. Néanmoins, il n'est pas possible de résoudre cette équation de façon analytique dans le cas de tissus biologiques. Le travail consistera à mettre en place d'autres approches.

Des méthodes de Monte Carlo pour la simulation de parcours de photons, reposant sur une représentation probabiliste de l'équation ERT, existent déjà dans la littérature (voir par exemple [WANG 95]). Cependant, le problème majeur de ces méthodes est qu'elles demandent la simulation d'un très grand nombre de photons et requiert donc des moyens informatiques importants.

Mon travail sera de mettre au point une stratégie alternative à ces méthodes de Monte Carlo, de démontrer la convergence et d'étudier la performance en terme de temps de calcul de ces nouveaux algorithmes. Cette nouvelle stratégie

consiste en une adaptation des algorithmes pour la synthèse d'image que l'on retrouvera dans [PAULY 00] et [VEACH 97], appelés *Metropolis Light Transport*, dans le cadre de la diffusion de la lumière dans les tissus.

Ce mémoire comportera quatre parties :

1. **Les chaînes de Markov :**

Une introduction nécessaire à la théorie des chaînes de Markov homogènes à temps discret. Dans cette partie, j'énoncerai les résultats classiques et en démontrerai une partie. Elle sera orientée bien entendu, vers les notions requises à la compréhension des algorithmes de Metropolis. J'introduirai tout d'abord les chaînes de Markov de façon générale, puis je ferai une partie concernant la classification des chaînes et des états afin de pouvoir définir les notions de récurrence, d'irréductibilité et d'ergodicité d'une chaîne. Je parlerai ensuite de mesures invariantes et réversibles puis démontrerai les théorèmes ergodiques sur lesquels reposent les algorithmes de Monte Carlo par Chaînes de Markov.

2. **Algorithmes de Monte Carlo par Chaînes de Markov :**

Ce chapitre sera une description des algorithmes MCMC pour le cas des chaînes de Markov à temps discret dans un espace d'états finis. Il sera orienté vers les algorithmes de Metropolis et comportera quelques exemples. Cette partie comprendra également une étude de la vitesse de convergence de ces algorithmes par une analyse spectrale.

3. **Algorithme de Metropolis pour la synthèse d'image :**

Cette partie reprendra en détail les algorithmes pour la synthèse d'image, d'une part en l'absence de milieu (propagation de la lumière dans le vide), où la lumière se déplace en ligne droite entre les frontières de la scène et est en partie absorbée et refractée à chaque instant où elle touche une surface. D'autre part en prenant en compte l'influence d'un milieu dans lequel la lumière peut être absorbée ou refractée à chaque instant.

4. **Adaptation au cadre de la propagation de la lumière dans le cerveau :**

Ce dernier chapitre présentera nos stratégies d'adaptation des algorithmes présentées dans le chapitre 3 au cas de la diffusion de la lumière dans le cerveau.

Chapitre 1

Chaînes de Markov

Dans ce chapitre, nous allons introduire les notions nécessaires à la compréhension des algorithmes de Métropolis mis en oeuvre dans les chapitres 3 et 4, en particulier les chaînes de Markov et les théorèmes dit ergodiques. L'ensemble des résultats présentés dans ce chapitre sont tirés de [GRIMMETT 01] et [LACROI 05]. Pour une revue plus complète de la question, nous réferrons donc à ces deux sources.

1.1 Généralités sur les chaînes de Markov

Définition 1.1.1. Soit E un ensemble quelconque. Un **processus aléatoire** X à valeur dans $(E, \mathcal{B}(E))$ est un quadruplet

$$X = (\Omega, \mathcal{F}, \{X_t\}_{t \in T}, \mathbb{P})$$

tel que $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité et, $\{X_t, t \in T\}$ une famille de variables aléatoires à valeur dans $(E, \mathcal{B}(E))$, indexées par un temps t .

Les X_t ne sont donc pas indépendantes. Dans la plupart des cas, $T = \{0, 1, 2, \dots\}$, et nous appelons alors le processus X un *processus à temps discret*. Dans d'autre cas, $T = \mathbb{R}$ ou $[0, +\infty)$ et X est alors appelé *processus à temps continu*.

Nous nous intéressons à présent aux processus à temps-discret. Dans le reste de ce mémoire, E désignera désormais un ensemble dénombrable et $\{X_0, X_1, \dots\}$ une famille de variables aléatoires à valeur dans $(E, \mathcal{P}(E))$. E est appelé **Espace des états**. Chaque X_n est donc une variable aléatoire (v.a.) discrète prenant N valeurs possibles, où $N = |E|$ (N peut éventuellement être infini).

Définition 1.1.2. Le processus aléatoire X est une **Chaîne de Markov**, s'il satisfait la condition de Markov :

$$\forall n \geq 1 \text{ et } \forall s, x_0, \dots, x_{n-1} \in E$$

$$\mathbb{P}(X_n = s | X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = x_0) = \mathbb{P}(X_n = s | X_{n-1} = x_{n-1})$$

Proposition 1.1.1. De façon équivalente, la condition de Markov se réécrit : $\forall n_0 < n_1 < \dots < n_k \leq n$,

$$\mathbb{P}(X_{n+1} = s | X_{n_k} = x_{n_k}, \dots, X_{n_1} = x_{n_1}, X_{n_0} = x_{n_0}) = \mathbb{P}(X_{n+1} = s | X_{n_k} = x_{n_k}).$$

Ainsi que :

$\forall m, n \geq 0,$

$$\mathbb{P}(X_{m+n} = s | X_m = x_m, \dots, X_1 = x_1, X_0 = x_0) = \mathbb{P}(X_{m+n} = s | X_m = x_m).$$

Exemple 1 : [GRIMMETT 01, p. 71]

Considérons le cas de la *marche aléatoire simple* : $S_n = S_0 + \sum_{i=1}^n X_i$ avec $(X_k)_{k=0,1,\dots,n}$ indépendants et identiquement distribués (i.i.d.) valant 1 avec probabilité p et -1 avec probabilité $q = 1 - p$ et S_0 étant fixé (par exemple $S_0 = 0$).

$$S_{n+1} = X_{n+1} + S_n$$

Conditionnellement à $S_n = X_0 + \dots + X_n$, la distribution de S_{n+1} dépend uniquement de la valeur prise par X_{n+1} , et ne dépend pas de S_0, \dots, S_{n-1} .

La marche aléatoire simple est donc une chaîne de Markov dans laquelle on prendra comme famille de variables aléatoires $(S_n)_{n=\{0,1,\dots\}}$ à valeur dans $(E, \mathcal{P}(E))$ où E est ici égal à \mathbb{Z} .

Remarques :

1. L'ensemble E étant supposé dénombrable, il peut être mis en bijection avec un sous-ensemble E' de \mathbb{N} . Ainsi sans perte de généralité, nous supposons dans la suite que $E = E'$
2. Pour tout état i de E , lorsque $X_n = i$ nous dirons que la chaîne est dans le $i^{\text{ème}}$ état (ou que la chaîne passe par la valeur i) à l'étape n .
3. Sachant la valeur prise par la chaîne à l'étape n on pourra caractériser l'étape $n+1$ de la chaîne grâce à ses **probabilités** (ou **noyaux**) de **transition** : $p_{ij,n} := \mathbb{P}(X_{n+1} = j | X_n = i)$, $\forall i, j \in E$.

Les noyaux de transition dépendent donc de trois paramètres : i, j, n . Dans ce mémoire, nous nous limiterons au cas des chaînes de Markov *homogènes*, c'est à dire que les noyaux de transition des chaînes ne dépendront plus de n :

Définition 1.1.3. 1. Une chaîne de Markov X est dite **homogène** lorsque la suite de ses probabilités de transition ne dépend pas de n , c'est à dire :

$$\forall n \geq 0, i, j \in E, \quad p_{ij,n} = p_{ij,0} := p_{ij}. \quad (1.1)$$

2. On appelle **matrice des transitions** P , la matrice de taille $N \times N$ des p_{ij} .

Proposition 1.1.2. P est une matrice **Stochastique**, c'est à dire :

- $p_{ij} \geq 0 \forall i, j \in E$
- $\sum_{j \in E} p_{ij} = 1 \forall i \in E$.

Remarques : Quelque soit l'étape n , la connaissance de la valeur prise par la chaîne à cette étape et la matrice des transitions P décrivent l'étape $n+1$ de la chaîne X . Pour décrire X sur un plus long-terme (étapes supérieures à $n+1$), nous avons recourt à la matrice des transition à l'étape n :

Définition 1.1.4. La matrice des transitions à l'étape n , notée $P(m, m+n)$ est la matrice des $p_{ij}(m, m+n) := \mathbb{P}(X_{m+n} = j | X_m = i)$

Remarques :

1. Sous l'hypothèse que X est homogène, on a pour tout $m \in \mathbb{N}$:

$$P(m, m+1) = P$$
2. Le théorème suivant montre que dans ce cas, $P(m, m+n)$ ne dépend pas de m :

Théorème 1.1.1. Equations de Chapman-Kolmogorov :

Soit X une chaîne de Markov homogène de matrice de transition P . Alors :

1.

$$p_{ij}(m, m+n+r) = \sum_{k \in E} p_{ik}(m, m+n) p_{kj}(m+n, m+n+r)$$

Ainsi :

$$P(m, m+n+r) = P(m, m+n) P(m+n, m+n+r)$$

2.

$$P(m, m+n) = P^n$$

Démonstration.

1.

$$\begin{aligned} p_{ij}(m, m+n+r) &= \mathbb{P}(X_{m+n+r} | X_m = i) \\ &= \sum_{k \in E} \mathbb{P}(X_{m+n+r} | X_{m+n} = k, X_m = i) \mathbb{P}(X_{m+n} = k | X_m = i) \\ &= \sum_{k \in E} \mathbb{P}(X_{m+n+r} | X_{m+n} = k) \mathbb{P}(X_{m+n} = k | X_m = i) \end{aligned}$$

2.

En utilisant le point 1 de ce théorème, on a :

$$\forall m \in \mathbb{N} \quad P(m, m+2) = \underbrace{P(m, m+1)}_P \underbrace{P(m+1, m+2)}_P = P^2$$

Hypothèse de récurrence : pour n fixé, $P(m, m+N) = P^N$.

Toujours d'après 1 :

$$P(m+N, m+N+1) = \underbrace{P(m, m+N)}_{P^N} \underbrace{P(m+N, m+N+1)}_P = P^{N+1}$$

Donc $\forall n \geq 2$, $P(m, m+n) = P^n$ □

Nous nous placerons par la suite uniquement dans le cadre des chaînes **homogènes**. Nous avons donc dans ce cas $p_{ij}(m, m+n) = p_{ij}(n)$ et les équations de Chapman-Kolmogorov se réécrivent :

$$p_{ij}(n+r) = \sum_{k \in E} p_{ik}(n) p_{kj}(r)$$

Soient i un état de E et $\mu_i^{(n)} = \mathbb{P}(X_n = i)$ les masses de X_n . Nous noterons $\mu^{(n)} = (\mu_i^{(n)}, i \in E)$

Lemme 1.1.1. On a pour tout entier n et m : $\mu^{(n+m)} = \mu^{(m)} P^n$.

Ainsi $\mu^{(n)} = \mu^{(0)} P^n$.

Démonstration.

$$\begin{aligned} \mu_j^{(n+m)} &= \mathbb{P}(X_{m+n} = j) = \sum_{i \in E} \underbrace{\mathbb{P}(X_{m+n} = j | X_m = i)}_{p_{ij}(n)} \underbrace{\mathbb{P}(X_m = i)}_{\mu_i^{(m)}} \\ &= \sum_{i \in E} \mu_i^{(m)} p_{ij}(n) = \left(\mu^{(m)} P^n \right)_j \end{aligned}$$

□

Nous obtenons donc le résultat important que l'évolution d'une chaîne de Markov homogène est caractérisée uniquement par sa matrice de transition P et sa loi initiale $\mu^{(0)}$, la loi de X_0 . Nous noterons plus simplement cette loi μ .

Proposition 1.1.3. X est une chaîne de Markov définie sur E , de matrice de transition P , si et seulement si :

Pour tout $i_0, \dots, i_n \in E$

$$\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mu_{i_0}^{(0)} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n}$$

Démonstration.

\Rightarrow) Supposons que X est une chaîne de Markov. Soit $A_k = \{X_k = i_k\}$.

$$\begin{aligned} \mathbb{P}(A_n \cap A_{n-1} \cap \dots \cap A_0) &= \mathbb{P}(A_n | A_{n-1} \cap A_{n-2} \cap \dots \cap A_0) \mathbb{P}(A_{n-1} \cap \dots \cap A_0) \\ \text{(X chaîne de Markov)} &= p_{i_{n-1}, i_n} \mathbb{P}(A_{n-1} \cap \dots \cap A_0) \\ &= \dots = p_{i_{n-1}, i_n} \dots p_{i_1, i_2} \mathbb{P}(A_1 \cap A_0) \\ &= \mu_{i_0}^{(0)} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n} \end{aligned}$$

\Leftarrow) Supposons que X vérifie :

$$\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mu_{i_0}^{(0)} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n} \quad \text{Pour tout } i_0, \dots, i_n \in E$$

Montrons que X est une chaîne de Markov de matrice de transition P .

Calculons :

$$\begin{aligned} \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0) &= \frac{\mathbb{P}(A_n | A_{n-1} \cap \dots \cap A_0)}{\mathbb{P}(A_{n-1} \cap \dots \cap A_0)} \\ &= \frac{\mu_{i_0}^{(0)} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n}}{\mu_{i_0}^{(0)} p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-2}, i_{n-1}}} = p_{i_{n-1}, i_n} \end{aligned}$$

D'où X vérifie la condition de Markov. □

Retour à l'exemple 1 :

Dans cet exemple, l'espace des états $E = \mathbb{Z}$ et ses probabilités de transition sont données par :

$$p_{ij} = \begin{cases} p & \text{si } j = i + 1 \\ q & \text{si } j = i - 1 \\ 0 & \text{sinon} \end{cases}$$

Pour toutes réalisations s_n de S_n nous avons $s_{n+1} - s_n = \pm 1$

Posons $r = |\{i; s_{i+1} - s_i = 1\}|$ et $l = |\{i; s_{i+1} - s_i = -1\}|$

Si nous avons $s_n = j$ et $s_0 = i$ alors $n = r + l$ et $j - i = r - l$ et donc

$$r = \frac{1}{2}(n + (j - i)) \text{ et } l = \frac{1}{2}(n - (j - i))$$

Ainsi :

$$p_{ij}(n) = \mathbb{P}(S_n = j | S_0 = i) = \begin{cases} \binom{n}{\frac{1}{2}(n + (j - i))} p^{\frac{1}{2}(n + (j - i))} q^{\frac{1}{2}(n - (j - i))} & \text{si } n + (j - 1) \text{ pair} \\ 0 & \text{sinon} \end{cases}$$

1.2 Classification des états

Définition 1.2.1. Un état $i \in E$ est dit **récurrent** (ou **persistant**) si :

$$\mathbb{P}(\exists n \geq 1 \text{ tel que } X_n = i | X_0 = i) = 1$$

Ce qui signifie que la probabilité d'un éventuel retour en i est égal à 1. Si cette probabilité est strictement inférieure à 1, l'état i est alors dit **transitoire** (ou **transient**). La chaîne est dite récurrente si tous ses états le sont.

Introduisons la probabilité $f_{ij}(n)$ que la chaîne X visite l'état j pour la première fois à l'étape n , sachant que X_0 vaut i :

$$f_{ij}(n) := \mathbb{P}\left(\bigcap_{k=1}^{n-1} \{X_k \neq j\}, \{X_n = j\} | X_0 = i\right) \text{ pour } n \geq 1 \text{ et } f_{ij}(0) := 0 \text{ pour tout } i, j \in E$$

de même posons f_{ij} la probabilité que la chaîne visite l'état j , sachant que X_0 vaut i :

$$f_{ij} = \sum_{n=1}^{+\infty} f_{ij}(n)$$

Par définition, nous avons que $f_{jj} = 1 \Leftrightarrow j$ est récurrent. Définissons les fonctions génératrices :

$$P_{ij}(s) = \sum_{n=0}^{+\infty} s^n p_{ij}(n) \qquad F_{ij}(s) = \sum_{n=0}^{+\infty} s^n f_{ij}(n)$$

Avec $p_{ij}(0)$ valant 1 si $i = j$ et 0. On a bien sur : $f_{ij} = F_{ij}(1)$

Théorème 1.2.1. Pour tout s et tout $i, j \in E$, on a :

- (a) $P_{ii}(s) = 1 + F_{ii}(s)P_{ii}(s)$
- (b) $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$ si $i \neq j$

Démonstration. Soit l'événement $B_m = \left\{ \bigcap_{k=1}^{m-1} X_k \neq j, X_m = j \right\}$. Définis de cette façon, les B_m sont disjoints, par conséquent on a :

$$\mathbb{P}(X_m = j | X_0 = i) = \sum_{k=1}^m \mathbb{P}(X_m = j \cap B_k | X_0 = i)$$

En utilisant la condition de Markov, on obtient :

$$\begin{aligned} \mathbb{P}(X_m = j \cap B_k | X_0 = i) &= \mathbb{P}(X_m = j | B_k, X_0 = i) \mathbb{P}(B_k = j | X_0 = i) \\ &= \mathbb{P}(X_m = j | X_k = j) \mathbb{P}(B_k = j | X_0 = i) \end{aligned}$$

D'où :

$$p_{ij}(m) = \sum_{k=1}^m f_{ij}(k) p_{jj}(m-k)$$

Par conséquent :

$$\begin{aligned}
\sum_{n \in \mathbb{N}^*} s^n p_{ij}(n) &= \sum_{n \in \mathbb{N}^*} \left(\sum_{k=1}^n f_{ij}(k) p_{jj}(n-k) \right) s^n \\
(\text{Par Toneli}) &= \sum_{k=1}^{+\infty} \sum_{n=k}^{+\infty} f_{ij}(k) p_{jj}(n-k) s^k s^{n-k} \\
&= \sum_{k=1}^{+\infty} f_{ij}(k) s^k \sum_{n=k}^{+\infty} p_{jj}(n-k) s^{n-k} \\
&= \sum_{k=1}^{+\infty} f_{ij}(k) s^k \sum_{n=0}^{+\infty} p_{jj}(n) s^n \\
&= F_{ij}(s) P_{jj}(s) \text{ car } f_{ij}(0) = 0
\end{aligned}$$

Or :

$$\sum_{n \in \mathbb{N}^*} s^n p_{ij}(n) = \sum_{n \in \mathbb{N}} s^n p_{ij}(n) - p_{ij}(0) = P_{ij}(s) - \delta_{ij}$$

D'où :

$$P_{ij}(s) - \delta_{ij} = F_{ij}(s) P_{jj}(s)$$

□

Corollaire 1.2.1.

1. L'état j est récurrent si $\sum_{n=0}^{+\infty} p_{jj}(n) = +\infty$. Dans ce cas $\sum_{n=0}^{+\infty} p_{ij}(n) = +\infty$, pour tout i tels que $f_{ij} > 0$
2. L'état j est transitoire si $\sum_{n=0}^{+\infty} p_{jj}(n) < +\infty$. Dans ce cas $\sum_{n=0}^{+\infty} p_{ij}(n) < +\infty$ pour tout i .

Corollaire 1.2.2. En particulier, ce résultat implique que si j est transitoire, alors $p_{ij} \rightarrow 0$ quand n tend vers l'infini et ce pour tout $i \in E$

Démonstration. D'après le théorème 1.2.1 (a), $P_{jj}(s) = [1 - F_{jj}(s)]^{-1}$ si $|s| < 1$. Donc $\lim_{s \rightarrow 1} P_{jj}(s) = +\infty \Leftrightarrow f_{jj} = F_{jj}(1) = 1 \Leftrightarrow j$ est récurrent.

Or $\lim_{s \rightarrow 1} P_{jj}(s) = \sum_{n \in \mathbb{N}} p_{jj}(n)$.

Par conséquent j est récurrent si et seulement si $\sum_{n \in \mathbb{N}} p_{jj}(n) = +\infty$

Enfin d'après 1.2.1 (b), si $\sum_{n \in \mathbb{N}} p_{jj}(n) = +\infty$ alors $\sum_{n \in \mathbb{N}} p_{ij}(n) = +\infty$ pour tout i tel que $f_{ij} > 0$ □

Posons T_j le temps du premier passage par j :

$$T_j = \min \{n \geq 1; X_n = j\},$$

avec $T_j = +\infty$ si ce passage n'a jamais lieu.

Définition 1.2.2.

1. Le temps de récurrence moyen est défini par :

$$\mu_i = \mathbb{E}[T_i | X_0 = i] = \begin{cases} \sum_{n=0}^{+\infty} n \mathbb{P}(T_i = n | X_0 = i) & \text{si } i \text{ récurrent,} \\ +\infty & \text{sinon.} \end{cases}$$

2. Si i est un état récurrent, il est appelé $\begin{cases} \text{nul} & \text{si } \mu_i = +\infty, \\ \text{positif} & \text{si } \mu_i < +\infty. \end{cases}$

Théorème 1.2.2. Un état récurrent i est nul si et seulement si

$$\lim_{n \rightarrow +\infty} p_{ii}(n) = 0.$$

Dans ce cas, $\lim_{n \rightarrow +\infty} p_{ij}(n) = 0 \forall j \in E$.

Définition 1.2.3.

1. La **période** $d(i)$ d'un état i est le plus grand diviseur commun parmi les entiers n tels que $p_{ii}(n) > 0$:

$$d(i) = \text{PGCD}\{n : p_{ii}(n) > 0\}.$$

2. i est dit **périodique** si $d(i) > 1$ et **apériodique** si $d(i) = 1$.
3. Un état est dit **ergodique** si il est **récurrent positif et apériodique**.
4. Une chaîne sera dite **ergodique** si tous ses états le sont.

Remarque :

Il est clair que si $p_{ii} > 0$ alors i est apériodique. D'une manière générale, il est facile de voir que :

i est apériodique si $\exists N \in \mathbb{N}; \mathbb{P}(X_n = i | X_0 = i) > 0 \forall n \geq N$.

1.3 Classification des chaînes

Définition 1.3.1. Soient X une chaîne de Markov de matrice de transition P et deux états $i, j \in E$.

1. On dit que i **conduit à j** , que l'on note $i \rightarrow j$ si :

$$\exists m \geq 0 \mathbb{P}(X_m = j | X_0 = i) > 0.$$

2. On dit que i et j **communiquent** que l'on note $i \leftrightarrow j$ si :

$$i \rightarrow j \text{ et } j \rightarrow i.$$

Propriété 1.3.1. La relation \leftrightarrow est une relation d'équivalence :

Elle est :

- (a) réflexive car $\mathbb{P}(X_0 = i | X_0 = i) = 1$ ($i \rightarrow i$),
- (b) symétrique par définition,
- (c) transitive :

Si $i \rightarrow j$ et $j \rightarrow k$ $\exists m, m' \geq 0$ tels que $\begin{cases} \mathbb{P}(X_m = j | X_0 = i) > 0 \\ \mathbb{P}(X_{m'} = k | X_0 = j) > 0 \end{cases}$

$$\begin{aligned} \mathbb{P}(X_{m+m'} = k | X_0 = i) &\geq \mathbb{P}(X_{m+m'} = k, X_m = j | X_0 = i) \\ &= \mathbb{P}(X_{m+m'} = k | X_0 = i, X_m = j) \mathbb{P}(X_m = j | X_0 = i) \\ &= \mathbb{P}(X_{m+m'} = k | X_m = j) \mathbb{P}(X_m = j | X_0 = i) \\ &= \mathbb{P}(X_{m'} = k | X_0 = j) \mathbb{P}(X_m = j | X_0 = i) > 0 \end{aligned}$$

Donc il existe $M = m + m'$ tel que $\mathbb{P}(X_M = k | X_0 = i) > 0$.

Théorème 1.3.1. Si $i \leftrightarrow j$, alors :

- a) i et j ont même période,
- b) i est transitoire $\Leftrightarrow j$ est transitoire,
- c) i est récurrent positif $\Leftrightarrow j$ est récurrent positif.

Démonstration.

a) Soient i et j deux états de E tels que $i \leftrightarrow j$, $d(i)$ et $d(j)$ les périodes respectives de i et j et soient $R(i) = \{n : p_{ii}(n) > 0\}$, et $R(j)$ défini de la même façon.

Comme $i \leftrightarrow j$, il existe n_1 et n_2 tel que $p_{ij}(n_1) > 0$ et $p_{ji}(n_2) > 0$. D'après les équations de Chapman Kolmogorov, on a :

$$p_{ii}(n_1 + n_2) = \sum_{k \in E} p_{ik}(n_1) p_{ki}(n_2) > p_{ij}(n_1) p_{ji}(n_2) > 0.$$

De plus pour tout $r \in R(j)$:

$$p_{ii}(n_1 + r + n_2) > p_{ij}(n_1) p_{jj}(r) p_{ji}(n_2) > 0.$$

Donc $n_1 + n_2 \in R(i)$ et $n_1 + r + n_2 \in R(i)$ donc $d(i)$ divise $n_1 + n_2$ et $n_1 + r + n_2$, il divise donc r . Par conséquent $d(i) \leq d(j)$.

On obtient l'égalité en échangeant le rôle de i et j .

b) Si $i \leftrightarrow j$ alors il existe $m, n \geq 0$ tels que :

$$\alpha = p_{ij}(m)p_{ji}(n) > 0.$$

D'après les équations de Chapman-Kolmogorov, on a pour $r \in \mathbb{N}$:

$$p_{ii}(m+r+n) \geq p_{ij}(m)p_{ji}(n+r) \geq p_{ij}(m)p_{ji}(n)p_{jj}(r) = \alpha p_{jj}(r).$$

En sommant de chaque cotés par rapport à r , on obtient donc :

$$\sum_{r \in \mathbb{N}} p_{jj}(r) < +\infty \text{ si } \sum_{r \in \mathbb{N}} p_{ii}(r) < +\infty.$$

Donc d'après le corollaire 1.2.1 (b), j est transitoire si i est transitoire. On obtient l'équivalence en échangeant le rôle de i et de j . \square

Définition 1.3.2. Soit C un ensemble d'états.

1. C est **fermé** si $p_{ij} = 0 \forall i \in C \text{ et } \forall j \notin C$,
2. C est **irréductible** si $\forall i, j \in C, i \leftrightarrow j$.

Théorème 1.3.2. Théorème de décomposition

Un espace des états E peut être partitionné de manière unique de la façon suivante :

$$E = T \cup C_1 \cup C_2 \dots$$

où T est un ensemble d'états transitoires, et les $C_k, k = 1, 2, \dots$ des ensembles d'états récurrents, fermés et irréductibles.

Démonstration. Soient C_1, C_2, \dots des ensembles irréductibles d'états récurrents (donc en particulier formant de classes d'équivalences pour la relation \leftrightarrow). Montrons que les C_k sont fermés. Soit C_r l'un de ces ensembles d'états et prenons i, j tels que $i \in C_r$ et $j \notin C_r$. Montrons que $p_{ij} = 0$. Comme C_r forme une classe d'équivalence, j ne conduit pas à i . Par conséquent :

$$\mathbb{P}(X_n \neq i \forall n \geq 1 | X_0 = i) \geq \mathbb{P}(X_1 = j | X_0 = i) = p_{ij} > 0.$$

Ce qui est contradictoire avec le fait que C_r un ensemble d'états récurrents puisque $\mathbb{P}(\exists n \text{ tel que } X_n = i | X_0 = i) = 1$ et donc $\mathbb{P}(X_n \neq i \forall n \geq 1 | X_0 = i) = 0$. \square

Lemme 1.3.1. Si E est fini, alors au moins un état est récurrent et tous les états récurrents sont positifs.

Démonstration. Si tous les états étaient transitoires, d'après le corollaire 1.2.2 on aurait pour tout $i \in E$:

$$\lim_{n \rightarrow +\infty} \sum_{j \in E} p_{ij}(n) = \sum_{j \in E} \lim_{n \rightarrow +\infty} p_{ij}(n) = 0,$$

ce qui est en contradiction avec le fait que $\sum_j p_{ij}(n) = 1$. \square

Lemme 1.3.2. Soit X une chaîne de Markov de matrice de transition P irréductible sur un espace d'état E fini. Alors d'après le théorème 1.3.1 et le lemme 1.3.1 tous les états de E sont récurrents positifs.

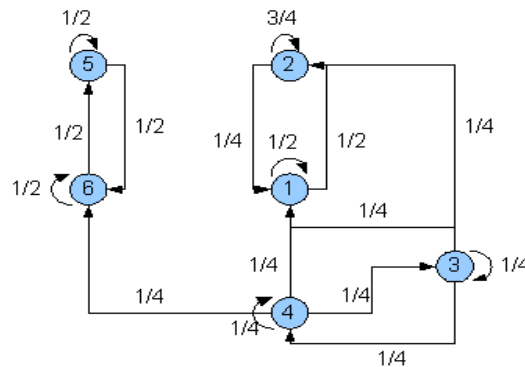


FIGURE 1.1 – Probabilités de transition des états

Exemple 2 :

Soit $E = \{1, 2, 3, 4, 5, 6\}$ et soit X une chaîne de Markov définie sur E avec matrice de transition P où

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

Nous avons représenté ces probabilités de transitions grâce au schéma de la figure précédente.

On peut voir que les ensembles $\{1, 2\}$ et $\{5, 6\}$ sont irréductibles et fermés car $1 \leftrightarrow 2$, $5 \leftrightarrow 6$ et que si à une étape n X_n est dans un de ces deux ensembles, la probabilité de quitter l'un de ces deux ensembles est nulle, donc les états 1, 2, 5, 6 sont récurrents positifs (positifs car E est fini). Les états 3 et 4 sont quant à eux transitoires car $3 \leftrightarrow 4 \rightarrow 6$, mais un retour depuis 6 vers 3 ou 4 est impossible. Tous les états sont apériodiques car pour tout $i \in E$, $p_{ii} > 0$. Par conséquent les états 1, 2, 5, 6 sont ergodiques.

1.4 Temps d'arrêt et Propriétés de Markov

Définition 1.4.1.

1. Nous appelons **temps aléatoire** T , toute variable aléatoire à valeur dans $\mathbb{N} \cup \{+\infty\}$.
2. Nous appelons **filtration**, toute suite croissante $\{\mathcal{F}_n\}$ de sous-tribus de $\mathcal{F} : \mathcal{F}_n \subset \mathcal{F}_{n+1}$. La filtration $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ (tribu engendrée par les X_0, \dots, X_n) est appelé **filtration naturelle** de X .
3. Soit T un temps aléatoire. T est un **temps d'arrêt** si l'événement $\{T \leq n\} \in \mathcal{F}_n$. Il n'est pas très difficile de voir que de façon équivalente, T est un temps d'arrêt si $\{T = n\} \in \mathcal{F}_n$.

Exemples :

Posons :

$$T_i = \inf\{n \geq 1; X_n = i\}.$$

Défini de cette façon, T_i est le temps d'arrêt désignant le temps de premier passage de la chaîne à l'état i à partir de l'étape 1 :

$$\{T_i = n\} = \{\inf\{k \geq 1; X_k = i\} = n\} = \underbrace{\{X_1 \neq i\}}_{\in \mathcal{F}_1} \cap \dots \cap \underbrace{\{X_{n-1} \neq i\}}_{\in \mathcal{F}_{n-1}} \cap \underbrace{\{X_n = i\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n.$$

De même, si l'on nomme pour $p \in \mathbb{N}^*$, $T_i^{(p)} = \inf\{n \geq T_i^{(p-1)}; X_n = i\}$ le p -ième temps de passage de la chaîne à l'état i où $T_i^{(0)} = T_i$, on peut montrer que $T_i^{(p)}$ est un temps d'arrêt.

En utilisant la notion de filtration, une définition alternative aux chaînes de Markov homogènes est :

Définition 1.4.2. Un processus X à valeur dans E est une chaîne de Markov (homogène) de matrice de transition P si pour toute fonction bornée de E dans \mathbb{R} et pour tout entier n , on a :

$$\mathbb{E}(f(X_{n+1})|\mathcal{F}_n) = Pf(X_n) := \sum_{j \in E} f(j)\mathbb{P}(X_{n+1} = j|X_n),$$

ou de façon équivalente, X est une chaîne de Markov si pour tout état j de E on a :

$$\mathbb{P}(X_{n+1} = j|\mathcal{F}_n) = \mathbb{P}(X_{n+1} = j|X_n).$$

Théorème 1.4.1. Propriété de Markov forte

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur E de matrice de transition P et de loi initiale μ , et soit T un temps d'arrêt.

Alors conditionnellement à l'événement $\{T < +\infty\}$, pour toute fonction $\phi \in C^b$ à valeur dans E , on a :

$$\mathbb{E}_\mu[\phi((X_{T+n})_{n \geq 0})|\mathcal{F}_T] = \mathbb{E}_{X_T}[\phi((X_n)_{n \geq 0})].$$

En d'autres termes, conditionnellement à l'événement $\{X_T = i_0, T < +\infty\}$, la chaîne $(X_{T+n})_{n \geq 0}$ est une chaîne de Markov de matrice P et de loi initiale

la loi δ_{i_0} (i.e. : $\mathbb{P}(X_0 = i_0) = 1$).

C'est à dire pour tout état j de E on a :

$$\mathbb{P}(X_{T+n} = j | \mathcal{F}_T \cap \{T < +\infty\}) = \mathbb{P}(X_{T+n} = j | X_T \cap \{T < +\infty\}).$$

De plus $(X_{T+n})_{n \geq 0}$ est indépendant des (X_0, \dots, X_{T-1}) .

Proposition 1.4.1. Soient E un espace d'états dénombrable, une fonction $f : E \rightarrow \mathbb{R}$ et j un état récurrent de E .

On pose :

$$Z_n = \sum_{k=T_j^{(n)}}^{T_j^{(n+1)}-1} f(X_k),$$

où

$$T_j^{(n+1)} = \inf\{n \geq T_j^{(n)}; X_{n+1} = j\} < +\infty \quad \mathbb{P}_j\text{-presque sûrement (} j \text{ récurrent)}$$

et désigne le temps du $(n+1)^{\text{eme}}$ retour en j .

Alors, conditionnellement à l'évènement $\{X_0 = j\}$ les Z_n sont i.i.d.

Démonstration. Soient $p \leq q$ deux entiers et montrons que pour toutes fonctions $\phi, \psi \in \mathcal{C}^b$, on a

$$\mathbb{E}_j[\phi(Z_p)\psi(Z_q)] = \mathbb{E}_j[\phi(Z_p)]\mathbb{E}_j[\psi(Z_q)].$$

Soient $\theta_n : \Omega \rightarrow \Omega$ une suite d'opérateurs de translation sur Ω , définies par $X_k \circ \theta_n = X_{n+k}$. On a par définition de Z_q et des θ_n :

$$\begin{aligned} \mathbb{E}_j[\phi(Z_p)\psi(Z_q)] &= \mathbb{E}_j\left(\phi(Z_p)\psi(Z_0 \circ \theta_{T_j^{(q)}})\right) \\ &= \mathbb{E}_j\left[\mathbb{E}_j\left(\phi(Z_p)\psi(Z_0 \circ \theta_{T_j^{(q)}}) | \mathcal{F}_{T_j^{(q)}}\right)\right]. \end{aligned}$$

Comme Z_p est $\mathcal{F}_{T_j^{(q)}}$ -mesurable (puisque $p \leq q$ et $Z_p = \sum_{k=T_j^{(p)}}^{T_j^{(p+1)}-1} f(X_k)$),

$$\mathbb{E}_j\left[\mathbb{E}_j\left(\phi(Z_p)\psi(Z_0 \circ \theta_{T_j^{(q)}}) | \mathcal{F}_{T_j^{(q)}}\right)\right] = \mathbb{E}_j\left[\phi(Z_p)E_j\left(\psi(Z_0 \circ \theta_{T_j^{(q)}}) | \mathcal{F}_{T_j^{(q)}}\right)\right].$$

En utilisant la propriété de Markov forte, on obtient alors :

$$\begin{aligned} \mathbb{E}_j[\phi(Z_p)\psi(Z_q)] &= \mathbb{E}_j\left[\phi(Z_p)E_{X_{T_j^{(q)}}}(\psi(Z_0))\right] \\ &= \mathbb{E}_j[\phi(Z_p)]E_j[\psi(Z_0)] \quad \forall \phi, \psi \in \mathcal{C}^b. \end{aligned}$$

En prenant $\phi = 1$, on obtient $E_j(\psi(Z_0)) = E_j(\psi(Z_q))$. Donc on a bien

$$\mathbb{E}_j[\phi(Z_p)\psi(Z_q)] = \mathbb{E}_j[\phi(Z_p)]\mathbb{E}_j[\psi(Z_q)] \quad \forall \phi, \psi \in \mathcal{C}^b.$$

Par conséquent Z_p et Z_q sont bien indépendants. \square

1.5 Invariance et réversibilité

Définition 1.5.1. Une *mesure* π sur un espace E dénombrable est un vecteur $(\pi_j, j \in E)$ dont tous les termes π_j sont positifs. De plus π est une **probabilité** si $\sum_{j \in E} \pi_j = 1$.

Définition 1.5.2. Une mesure est dite *invariante* ou *stationnaire* pour la chaîne définie pour la matrice de transition P si :

$$\pi P = \pi,$$

c'est à dire $\pi_j = \sum_{i \in E} \pi_i p_{ij} \quad \forall j \in E$.

Remarque :

Si π est invariante pour P , nous avons également :

$$\pi P^2 = (\pi P)P = \pi P = \pi.$$

Ainsi par récurrence on a que $\pi P^n = \pi$.

Si la loi π de X_0 est invariante pour P , alors toutes les variables X_n ont aussi π pour loi. La loi des X_n est donc invariante au cours du temps.

Résoudre le système d'équations linéaires afin de trouver la mesure invariante n'est pas toujours aisé. En général, on prouvera une condition plus forte mais plus facile à établir pour la mesure π qui est la condition de *réversibilité*.

Considérons une chaîne de Markov $\{X_n : 0 \leq n \leq N\}$, irréductible et récurrente positive, de matrice de transition P et de distribution stationnaire π ; Supposons de plus que X_n admette π pour distribution quelque soit n . Nous définissons la chaîne 'renversée' Y par $Y_n = X_{N-n}$.

Théorème 1.5.1. Y est une chaîne de Markov, de matrice de transition ayant pour éléments :

$$\mathbb{P}(Y_{n+1} = j | Y_n = i) = (\pi_j / \pi_i) p_{ji}$$

Démonstration.

$$\begin{aligned} \mathbb{P}(Y_{n+1} = i_{n+1} | Y_n = i_n, \dots, Y_0 = i_0) &= \frac{\mathbb{P}\left(\bigcap_{k=0}^{n+1} \{Y_k = i_k\}\right)}{\mathbb{P}\left(\bigcap_{k=0}^n \{Y_k = i_k\}\right)} \\ &= \frac{\mathbb{P}(X_{N-n-1} = i_{n+1}, X_{N-n} = i_n, \dots, X_N = i_0)}{\mathbb{P}(X_{N-n} = i_n, \dots, X_N = i_0)} \\ &= \frac{\pi_{i_{n+1}} p_{i_{n+1}, i_n} p_{i_n, i_{n-1}} \cdots p_{i_1, i_0}}{\pi_{i_n} p_{i_n, i_{n-1}} \cdots p_{i_1, i_0}} \\ &= \frac{\pi_{i_{n+1}}}{\pi_{i_n}} p_{i_{n+1}, i_n} \end{aligned}$$

□

Définition 1.5.3. *Sous les hypothèses précédentes pour X , on dit que X est réversible si la matrice de transition de X est la même que celle de Y . Ce qui revient à dire que X est réversible si :*

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j \in E.$$

Proposition 1.5.1. *X est réversible pour $\pi \Rightarrow \pi$ est une mesure invariante.*

Démonstration. $\pi_i p_{ij} = \pi_j p_{ji} \Rightarrow \sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$
Ainsi $\pi P = \pi$ □

Retour à l'exemple 2 :

Prenons l'ensemble fermé et irréductible $C = \{1, 2\}$ dans le cas où $X_0 \in C$.
Considérons la matrice de transition associée à X :

$$P_C = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

Comme X est irréductible et que tous ses états sont récurrents positifs, elle admet une unique distribution invariante (on le verra dans la section suivante). On peut voir facilement qu'elle est donnée par $\pi = (1/3 \quad 2/3)$ (On a bien $\pi P_C = \pi$). Par conséquent ses temps de récurrence moyens sont donnés par $\mathbb{E}(T_1 | X_0 = 1) = \pi_1^{-1} = 3$ et $\mathbb{E}(T_2 | X_0 = 2) = \pi_2^{-1} = \frac{3}{2}$

De la même façon, on peut calculer $\mathbb{E}(T_5 | X_0 = 5)$ et $\mathbb{E}(T_6 | X_0 = 6)$

1.6 Théorèmes ergodiques

Avant d'énoncer le premier résultat, qui n'est rien d'autre qu'une loi forte des grands nombres adaptées aux chaînes de Markov irréductibles, rappelons la loi forte des grands nombres originale, sur laquelle la démonstration de ce premier théorème ergodique repose.

Théorème 1.6.1. Loi Forte des Grands Nombres :

Soient X une variable aléatoire telle que $\mathbb{E}(|X|) < +\infty$ et (X_1, \dots, X_n) n variables aléatoires indépendantes et identiquement distribuées de même loi que X . Alors :

$$\lim_{n \rightarrow +\infty} \frac{X_1 + \dots + X_n}{n} = \mathbb{E}(X)$$

Théorème 1.6.2. Une chaîne X de matrice de transition irréductible P admet une distribution invariante π si et seulement si tous ses états sont récurrents positifs. Dans ce cas :

1. π est l'unique distribution stationnaire et est donnée par :

$$\pi_i = \frac{1}{\mathbb{E}(T_i | X_0 = i)} \quad \forall i \in E.$$

2. Pour toute loi initiale μ et toute fonction $f : E \rightarrow \mathbb{R}$ telle que

$$\sum_{j \in E} |f(j)| \pi_j < +\infty, \text{ on a :}$$

$$\boxed{\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow +\infty]{} \sum_{j \in E} f(j) \pi_j \quad \mathbb{P}_\mu\text{-p.s.}}$$

Démonstration.

1. Supposons tout d'abord que π est une distribution stationnaire de la chaîne supposée irréductible. Comme π stationnaire, on a : $\pi_j = \sum_i \pi_i p_{ij}(n)$. Or si tous les états étaient transitoires, on aurait $p_{ij}(n) \rightarrow 0 \quad \forall i, j \in E$, par conséquent

$$\pi_j = \sum_{i \in E} \pi_i p_{ij}(n) \rightarrow 0 \text{ quand } n \rightarrow +\infty \quad \forall i, j \in E.$$

En effet, soit F un sous-ensemble fini de E , on a :

$$\begin{aligned} \sum_{i \in E} \pi_i p_{ij}(n) &\leq \sum_{i \in F} \pi_i p_{ij}(n) + \sum_{i \in E-F} \pi_i \text{ car } p_{ij}(n) \leq 1 \\ &\rightarrow \sum_{i \in E-F} \pi_i \text{ quand } n \rightarrow +\infty \\ &\rightarrow 0 \text{ quand } F \uparrow E \text{ (} F \text{ tend vers } E \text{)}. \end{aligned}$$

Ce qui est contradictoire avec le fait que $\sum_{i \in E} \pi_i = 1$. Donc tous les états sont récurrents (la chaîne étant irréductible).

Montrons à présent que si π existe, alors tous les états sont récurrents positifs et que $\pi_i = \mu_i^{-1}$ pour tout i de E ,

où $\mu_i = \mathbb{E}(T_i | X_0 = i)$, et $T_i = \inf\{n \geq 1, X_n = i\}$.

Supposons que X_0 a pour distribution π , donc que $\mathbb{P}(X_0 = i) = \pi_i$ pour tout i .

On a :

$$\begin{aligned} \pi_j \mu_j &= \mathbb{E}(T_j | X_0 = j) \mathbb{P}(X_0 = j) \\ &= \sum_{n=1}^{+\infty} \mathbb{P}(T_j \geq n | X_0 = j) \mathbb{P}(X_0 = j) \\ &= \sum_{n=1}^{+\infty} \mathbb{P}(T_j \geq n, X_0 = j). \end{aligned}$$

Or on a :

$$\mathbb{P}(T_j \geq 1, X_0 = j) = \mathbb{P}(X_0 = j).$$

Et pour tout $n \geq 2$,

$$\mathbb{P}(T_j \geq n, X_0 = j) = \mathbb{P}(X_0 = j, X_m \neq j \forall 1 \leq m \leq n-1).$$

Comme :

$$\mathbb{P}(X_m \neq j \forall 1 \leq m \leq n-1) = \mathbb{P}(X_0 = j, X_m \neq j \forall 1 \leq m \leq n-1) + \mathbb{P}(X_0 \neq j, X_m \neq j \forall 1 \leq m \leq n-1),$$

on a :

$$\begin{aligned} \mathbb{P}(X_0 = j, X_m \neq j \forall 1 \leq m \leq n-1) &= \mathbb{P}(X_m \neq j \forall 1 \leq m \leq n-1) - \mathbb{P}(X_m \neq j \forall 0 \leq m \leq n-1) \\ (\text{X homogène}) &= \mathbb{P}(X_m \neq j \forall 0 \leq m \leq n-2) - \mathbb{P}(X_m \neq j \forall 0 \leq m \leq n-1) \\ &= a_{n-2} - a_{n-1} \end{aligned}$$

où $a_n := \mathbb{P}(X_m \neq j \forall 0 \leq m \leq n)$.

On a donc :

$$\begin{aligned} \pi_j \mu_j &= \mathbb{P}(X_0 = j) + \lim_{R \rightarrow +\infty} \underbrace{\sum_{n=2}^R (a_{n-2} - a_{n-1})}_{a_0 - a_{R-1}} \\ &= \mathbb{P}(X_0 = j) + \mathbb{P}(X_0 \neq j) - \lim_{n \rightarrow +\infty} a_n = 1 - \lim_{n \rightarrow +\infty} a_n. \end{aligned}$$

Or comme j est récurrent, $\lim_{n \rightarrow +\infty} a_n = \mathbb{P}(X_m \neq j \forall m \geq 0) = 0$.

Par conséquent $\pi_j \mu_j = 1$ Montrons à présent que π charge tous les points de E (i.e. : $\pi_i > 0$ pour tout $i \in E$).

Supposons au contraire qu'il existe un état j pour lequel $\pi_j = 0$. Alors :

$$0 = \pi_j = \sum_{k \in E} \pi_k p_{kj}(n) \geq \pi_i p_{ij} \quad \forall i \in E, n \in \mathbb{N}.$$

X étant irréductible, $i \rightarrow j$ pour tout $i, j \in E$. Ainsi il existe n tel que $p_{ij}(n) > 0$, ce qui implique que $\pi_i = 0$, et donc par irréductibilité de la chaîne, $p_i = 0$ pour tout i . Ce qui est en contradiction avec $\sum_i \pi_i = 1$. Donc $\pi_j > 0$ pour tout état j de E . On peut alors écrire :

$$\mu_i = \frac{1}{\pi_i} < +\infty$$

Ce qui prouve également l'unicité de la mesure invariante.

Ainsi, si π existe, elle est unique et tous les états de la chaîne sont récurrents positifs.

On pourrait montrer réciproquement que si tous les états de la chaîne sont récurrents positifs, alors la distribution invariante existe et est donnée par

$$\pi_i = \frac{1}{\mu_i}.$$

2. Nous reprenons pour ce point une démonstration que l'on pourra retrouver dans [LACROI 05].

Définissons la mesure λ^x sur E par $\lambda^x = (\lambda_j^x; j \in E)$ où :

$$\lambda_j^x = \mathbb{E}_x \left[\sum_{k=0}^{T_x-1} \mathbb{1}(X_k = j) \right]$$

On a également :

$$\begin{aligned} \lambda^x(f) &= \sum_{j \in E} f(j) \lambda_j^x \\ &= \mathbb{E}_x \left[\sum_{k=0}^{T_x-1} \sum_{j \in E} f(j) \mathbb{1}(X_k = j) \right] \\ &= \mathbb{E}_x \left[\sum_{k=0}^{T_x-1} f(X_k) \right]. \end{aligned}$$

Remarquons que $\lambda_x^x = 1$ et $\lambda^x \left(\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right) = \mathbb{E}_x(T_x)$.

Nous allons montrer tout d'abord le résultat plus général suivant : pour tout $f, g \in l^1(\lambda^x)$, tel que $\lambda^x(g) \neq 0$, on a :

$$\mathbb{P}_x \left(\lim_{n \rightarrow +\infty} \frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} = \frac{\lambda^x(f)}{\lambda^x(g)} \right) = 1.$$

Prenons f positive et posons :

$$Z_p = \sum_{k=T_x^{(p)}}^{T_x^{(p+1)}-1} f(X_k).$$

D'après la proposition 1.4.1, tous les états sont récurrents donc les $T_x^{(p)}$ sont finis presque sûrement quelque soit $x \in E$. Ainsi les Z_p , $p = 0, \dots, n-1$ sont i.i.d, par conséquent d'après la loi forte des grands nombres, on a :

$$\frac{Z_0 + \dots + Z_{n-1}}{n} = \frac{1}{n} \sum_{k=0}^{T_x^{(n)}-1} f(X_k) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}[Z_0] = \lambda^x(f).$$

Soit $\nu(m)$ le nombre de retour en x de la chaîne X avant le temps m . Ainsi,

$$\nu(m) = \sum_{k=1}^{m-1} \mathbb{1}(X_k = x) \text{ et donc conditionnellement à } X_0 = x,$$

$$\text{on a : } \nu(m) + 1 = \sum_{k=0}^{m-1} \mathbb{1}(X_k = x). \text{ De plus :}$$

$$T_x^{(\nu(m))} < m \leq T_x^{(\nu(m)+1)}$$

Par conséquent :

$$\frac{\nu(m)}{\nu(m) + 1} \frac{\sum_{k=0}^{T_x^{(\nu(m))}-1} f(X_k)}{\nu(m)} \leq \frac{\sum_{k=0}^{m-1} f(X_k)}{\sum_{k=0}^{m-1} \mathbb{1}(X_k = x)} \leq \frac{\sum_{k=0}^{T_x^{(\nu(m)+1)}-1} f(X_k)}{\nu(m) + 1}$$

Comme $\nu(m)$ tend vers l'infini quand m tend vers l'infini, le terme de gauche et de droite tendent chacun vers $\lambda^x(f)$ presque sûrement. On a donc :

$$\frac{\sum_{k=0}^{m-1} f(X_k)}{\sum_{k=0}^{m-1} \mathbb{1}(X_k = x)} \xrightarrow{m \rightarrow +\infty} \lambda^x(f).$$

Ensuite, en écrivant $f = f^+ - f^-$, on voit que cette dernière limite est vérifiée pour tout $f \in l^1(\lambda^x)$; et si $f, g \in l^1(\lambda^x)$, en faisant le rapport on obtient bien le résultat annoncé.

Si l'on prouve que la mesure λ^x est invariante, grâce à l'unicité (à constante près) d'une telle mesure (d'après 1.), en prenant $g \equiv 1$ on aura bien le résultat annoncé.

Montrons donc que pour tout $f \in l^1(\lambda^x)$, on a bien $(\lambda^x P)(f) = \lambda^x(f)$:

$$\begin{aligned} (\lambda^x P)f = \lambda^x(Pf) &= \mathbb{E}_x \left[\sum_{k=0}^{T_x-1} Pf(X_k) \right] \\ &= \sum_{k \geq 0} \mathbb{E}_x [\mathbb{1}(k < T_x) Pf(X_k)]. \end{aligned}$$

Or par définition,

$$Pf(X_k) = \mathbb{E}_x[f(X_{k+1}) | \mathcal{F}_k].$$

Par conséquent :

$$\begin{aligned} (\lambda^x P)f &= \sum_{k \geq 0} \mathbb{E}_x [\mathbb{1}(k < T_x) \mathbb{E}_x(f(X_{k+1}) | \mathcal{F}_k)] \\ &= \sum_{k \geq 0} \mathbb{E}_x [\mathbb{E}_x(\mathbb{1}(k < T_x) f(X_{k+1}) | \mathcal{F}_k)] \text{ car } \mathbb{1}(k < T_x) \text{ est } \mathcal{F}_k\text{-mesurable} \\ &= \sum_{k \geq 0} \mathbb{E}_x [\mathbb{1}(k < T_x) f(X_{k+1})] \\ &= \mathbb{E}_x \left[\sum_{k=0}^{T_x-1} f(X_{k+1}) \right] \\ &= \mathbb{E}_x \left[\sum_{k=1}^{T_x} f(X_k) \right] \\ &= \lambda^x(f) - \mathbb{E}_x[f(X_0)] + \mathbb{E}_x[f(X_{T_x})] \\ &= \lambda^x(f) - f(x) + f(x) = \lambda^x(f). \end{aligned}$$

□

Théorème 1.6.3. (Théorème ergodique)

Soit X une chaîne de Markov de matrice de transition P irréductible et apériodique.

Alors :

$$\lim_{n \rightarrow +\infty} p_{ij}(n) = \frac{1}{\mathbb{E}(T_j | X_0 = j)} \quad \forall i, j \in E$$

Remarques :

Sous les hypothèses du théorème précédent :

1. Si X est transitoire ou récurrente nul, alors $p_{ij}(n) \rightarrow 0 \quad \forall i, j$
2. Si la chaîne est récurrente positive, alors $p_{ij}(n) \rightarrow \pi_j \quad \forall i, j$ où π est l'unique distribution stationnaire ; c'est à dire $\mathcal{L}(X_n | X_0) \rightarrow \pi$
3. on peut voir que

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = j) = \lim_{n \rightarrow +\infty} \sum_i p_{ij}(n) \mathbb{P}(X_0 = i) = \frac{1}{\mathbb{E}(T_j | X_0 = j)}.$$

D'où, si la chaîne est récurrente positive, $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = j) = \pi_j$, et ce, quelque soit la distribution de X_0 ; c'est à dire $\mathcal{L}(X_n) \rightarrow \pi$.

Pour la démonstration de ce théorème, nous allons avoir recours à deux lemmes :

Lemme 1.6.1. Si X est une chaîne de Markov irréductible et apériodique, alors

$$\forall (i, j) \in E^2 \quad \exists N \text{ tel que } p_{ij}(n) > 0 \quad \forall n \geq N$$

Lemme 1.6.2. Si X est une chaîne de Markov sur E récurrente de matrice de transition P irréductible, alors tous les points de E sont visités une infinité de fois :

$$\forall i, j \in E, \quad \mathbb{P}(X_n = y \text{ pour une infinité de } n | X_0 = x) = 1$$

Ainsi $\mathbb{P}(T_j < +\infty | X_0 = i) = 1$ pour tout $i, j \in E$

Pour ce théorème, nous ne démontrerons pas le cas où X est récurrente nulle. (une démonstration peut être trouvée dans [GRIMMETT 01]).

Démonstration. (Théorème 1.6.3)

Si la chaîne est transitoire, alors le résultat découle directement du corollaire 1.2.2. Si la chaîne est récurrente, on utilise alors la technique appelée coupling : On construit un couple de chaînes $Z = (X, Y)$, X et Y étant deux chaînes de Markov *indépendantes* sur E , de même matrice de transition P . On remarque facilement que Z est une chaîne de Markov de matrice de transition :

$$\begin{aligned} p_{ij,kl} &= \mathbb{P}(Z_{n+1} = (k, l) | Z_n = (i, j)) \\ &= \mathbb{P}(X_{n+1} = k | X_n = i) \mathbb{P}(Y_{n+1} = l | Y_n = j) \text{ par indépendance de } X \text{ et } Y \\ &= p_{ik} p_{jl} \end{aligned}$$

Comme X est irréductible et apériodique, d'après le lemme 1.6.1, il existe N tel que :

$$p_{ij,kl}(n) = p_{ik}(n)p_{jl}(n) > 0 \text{ pour tout } n \geq N.$$

En particulier, Z est donc irréductible.

Supposons à présent que X est récurrente positive. Alors, d'après le théorème 1.6.2, il existe une unique distribution stationnaire et on peut voir que Z admet également une distribution stationnaire donnée par $\nu_{ij} = \pi_i \pi_j$:

$$\sum_i \sum_j \nu_{ij} p_{ij,kl} = \sum_i \sum_j \pi_i \pi_j p_{ik} p_{jl} = \sum_i \pi_i p_{ik} \sum_j \pi_j p_{jl} = \pi_k \pi_l = \nu_{kl}$$

Par conséquent, Z est récurrente positive, d'après le théorème précédent. Posons à présent l'ensemble :

$$T = \min\{n \geq 1, Z_n = (s, s) \text{ quelque soit } s \in E\}.$$

T est donc le premier instant où $X_n = Y_n$.

D'après le lemme 1.6.2, T est donc fini presque sûrement. De plus, remarquons que si $X_m = Y_m$, alors quelque soit l'entier $n \geq m$, les variables aléatoires X_n et Y_n sont identiquement distribuées, et leurs distributions ne dépendent que de P et de la valeur des chaînes à l'étape m .

Ainsi sous la condition $\{T \leq n\}$, X_n et Y_n ont la même distribution. Nous pouvons donc écrire :

$$\begin{aligned} p_{ik}(n) &= \mathbb{P}(X_n = k | X_0 = i, Y_0 = j) \text{ puisque } X \text{ et } Y \text{ sont indépendantes} \\ &= \mathbb{P}(X_n = k, T \leq n | Z_0 = (i, j)) + \mathbb{P}(X_n = k, T > n | Z_0 = (i, j)) \\ &= \mathbb{P}(Y_n = k, T \leq n | Z_0 = (i, j)) + \mathbb{P}(X_n = k, T > n | Z_0 = (i, j)) \text{ car } Y_n \text{ et } X_n \text{ ont la même loi} \\ &\leq \mathbb{P}(Y_n = k | Z_0 = (i, j)) + \mathbb{P}(T > n | Z_0 = (i, j)) \end{aligned}$$

Or, $\mathbb{P}(Y_n = k | Z_0 = (i, j)) + \mathbb{P}(T > n | Z_0 = (i, j)) = p_{jk}(n) + \mathbb{P}(T > n | Z_0 = (i, j))$.

On permutant les rôles de i et j , on obtient :

$$|p_{ik}(n) - p_{jk}(n)| \leq \mathbb{P}(T > n | Z_0 = (i, j)) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

Puisque T est fini presque sûrement. Ainsi :

$$\lim_{n \rightarrow +\infty} p_{ik}(n) - p_{jk}(n) = 0 \text{ pour tout } i, j, k \in E.$$

Ce qui montre que si la limite de $p_{ik}(n)$ existe, elle ne dépend pas de i , et donc $p_{ik}(n) \rightarrow \pi_k$.

Montrons que cette limite existe. Nous pouvons écrire :

$$p_{ik}(n) - \pi_k = \sum_j \pi_j (p_{ik}(n) - p_{jk}(n)) \text{ par invariance de } \pi.$$

Il suffit alors de montrer que cette somme tend vers 0. Soit donc F un sous ensemble fini de E .

$$\begin{aligned} \sum_j \pi_j |p_{ik}(n) - p_{jk}(n)| &\leq \sum_{j \in F} |p_{ik}(n) - p_{jk}(n)| + 2 \sum_{j \in E-F} \pi_j \\ &\rightarrow 2 \sum_{j \in E-F} \pi_j \text{ quand } n \rightarrow +\infty \\ &\rightarrow 0 \text{ quand } F \uparrow E. \end{aligned}$$

□

L'application directe de ces théorèmes ergodiques, et que si on peut simuler une chaîne de markov ergodique $(X_k)_{k=0,1,\dots}$ de matrice de transition irréductible et admettant π pour unique mesure invariante, alors :

1. Pour un n assez grand et pour toute fonction appartenant à $l^1(\pi)$, pour une variable aléatoire X de loi π , l'esperance $\mathbb{E}(f(X))$ peut être approché par la moyenne empirique $\frac{1}{n} \sum_{k=1}^n f(X_k)$.
2. À partir d'un certain rang N , les lois de X_N, X_{N+1}, \dots seront proches de la loi de π

Le chapitre suivant décrit donc des algorithmes permettant de simuler de telles chaînes de Markov à partir d'une mesure π .

Chapitre 2

Méthode de Monte-Carlo par Chaîne de Markov

2.1 Introduction

Les méthodes de Monte Carlo sont des algorithmes visant à approximer des intégrales, en utilisant la notion d'espérance probabiliste. Par exemple, si l'on cherche à évaluer la valeur de $I = \int_{\mathbb{R}} h(x)e^{-x} dx$, pour une certaine fonction h , l'idée sera de simuler dans ce cas N réalisations $\{x_1, \dots, x_n\}$ indépendantes d'une variable aléatoire de loi exponentielle $\mathcal{E}(1)$. Ainsi, en vertu de la loi forte des grands nombres, un estimateur consistant et sans biais de I sera donné par

$$\hat{I} = \frac{h(X_1) + \dots + h(X_N)}{N}.$$

Cependant, dans le cas d'intégrales plus complexes, comme par exemple $\int_D h(x)\pi(x)dx$ où π peut être mis sous la forme $\pi(x) = Z^{-1}\exp(-V(x))$ pour une fonction V pouvant être compliquée à expliciter (voir par exemple le cas des intégrales de Gibbs que l'on retrouvera dans [BENAIM 05]), la *constante de normalisation* $Z = \int_{\mathbb{R}} \exp(-V(x))dx$ est en général très difficile à évaluer et on ne pourra pas par conséquent utiliser les algorithmes de Monte Carlo classiques pour simuler π . Les algorithmes appelés MCMC (**M**onte **C**arlo **M**arkov **C**hain) prennent alors ici un grand intérêt (en plus de leurs faibles temps de simulation), car la connaissance de cette constante de normalisation n'est pas nécessaire pour la mise en oeuvre de l'algorithme. En contre-partie, ce dernier produira un échantillon non-indépendant, qui sera une chaîne de Markov de noyau de transition irréductible et admettant π pour distribution invariante. Si l'on peut montrer que la chaîne produite est ergodique, la convergence de \hat{I} vers l'intégrale que l'on cherche à évaluer sera assurée par le théorème 1.6.2 du chapitre 1. De plus, pour un nombre N suffisamment grand de simulations, la loi de X_N sera proche de π . En d'autres termes, après un temps n_0 ("burn-in period") les $x_{n_0+1}, x_{n_0+2}, \dots$ produits par la chaîne pourrons approximativement être vu comme des simulations de la loi cible π .

Dans ce chapitre nous nous placerons dans le cadre où E est un espace d'états fini et nous considérerons π une loi sur E connue à constante multiplicative près,

que l'on cherche à simuler.

Dans un premier temps nous présenterons l'algorithme de Metropolis initial, appelé également *algorithme de Metropolis indépendant*. Dans une seconde partie nous exposerons l'algorithme général, en donnant des preuves concernant la convergence de cet algorithme vers la loi cible. Enfin, après quelques exemples, nous nous intéresserons à la vitesse de convergence de ces algorithmes grâce à une analyse spectrale.

2.2 Algorithme de Metropolis indépendant

L'algorithme de Metropolis est un algorithme permettant de produire une chaîne de Markov réversible par rapport à la loi π , connue à constante multiplicative près, que l'on cherche à simuler.

On se donne une matrice de transition markovienne $Q = (q_{ij})_{i,j}$ sur E , appelé *matrice de sélection* (en général, de telle sorte que l'algorithme soit simple à mettre en oeuvre) qui soit irréductible et symétrique :

$$\forall (i, j) \in E^2, q_{ij} = q_{ji}.$$

L'algorithme est le suivant :

Algorithme :

Etape 0 : Initialiser $X_0 = i_0$, X_0 de loi quelconque.

Etape n+1 :

$$\left\{ \begin{array}{l} \text{On tire } i_{n+1} \text{ dans la loi } Q(i_n, \cdot) = (q_{i_n j}, j \in E) \\ \text{On tire } u \text{ aléatoirement dans } [0, 1] \\ \text{Si } u < \frac{\pi_{i_{n+1}}}{\pi_{i_n}}, \text{ on } \mathbf{accepte} \text{ la selection : } X_{n+1} = i_{n+1} \\ \text{Si } u > \frac{\pi_{i_{n+1}}}{\pi_{i_n}}, \text{ on } \mathbf{refuse} \text{ la selection : } X_{n+1} = i_n \end{array} \right.$$

2.3 Algorithme de Metropolis généralisé

Q n'est ici pas nécessairement symétrique, mais doit vérifier :

$$\forall (x, y) \in E^2, q_{ij} > 0 \Rightarrow q_{ji} > 0.$$

On se donne également une fonction (*acceptance function*) $h :]0, +\infty[\rightarrow]0, 1]$, vérifiant :

$$h(u) = uh\left(\frac{1}{u}\right) \text{ quelque soit } u.$$

Par Exemple :

- $h(u) = \min(1, u)$ (fonction d'acceptation de Metropolis et Hasting)

- $h(u) = \frac{u}{1+u}$ (fonction d'acceptation de Barker)

Enfin pour $i, j \in E$, $i \neq j$ posons :

$$\alpha(i, j) = \begin{cases} h\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) & \text{si } q_{ij} \neq 0 \\ 0 & \text{sinon} \end{cases}$$

On construit alors une matrice de transition P définie par :

$$p_{ij} = \begin{cases} q_{ij} \alpha(i, j) & \text{si } i \neq j \\ 1 - \sum_{k \neq i} p_{ik} & \text{si } i = j \end{cases}$$

Proposition 2.3.1. *Supposons que π charge tous les points de E . La matrice P est réversible par rapport à π (donc π invariante par rapport à P). Elle est irréductible si Q est choisie irréductible. Si de plus $h(u) < 1$ (Par exemple $h(u) = \frac{u}{1+u}$), elle est apériodique.*

Démonstration. Montrons que π est réversible pour P , c'est à dire pour tout $(x, y) \in E^2$:

$$\pi_i p_{ij} = \pi_j p_{ji}.$$

Pour $i = j$, le résultat est trivial. Pour $i \neq j$:

$$\begin{aligned} \pi_i p_{ij} &= \pi_i q_{ij} \alpha(i, j) \\ &= \pi_i q_{ij} h\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) \\ &= \pi_j q_{ji} \frac{\pi_i q_{ij}}{\pi_j q_{ji}} h\left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}}\right) \\ &= \pi_j q_{ji} h\left(\frac{\pi_i q_{ij}}{\pi_j q_{ji}}\right) \text{ car } h(u) = uh\left(\frac{1}{u}\right) \\ &= \pi_j q_{ji} \alpha(j, i) = \pi_j p_{ji}, \end{aligned}$$

Donc π est bien réversible pour P .

Si Q est irréductible $\forall (i, j) \in E^2$ il existe $n \in \mathbb{N}$ tel que $q_{ij}(n) > 0$. h étant strictement positive, $p_{ij}(n) > 0$ donc P est irréductible. \square

L'algorithme suivant (Métropolis) correspond à la simulation de la chaîne de Markov de noyau P :

Algorithme :

Étape 0 : Initialiser $X_0 = i_0$, X_0 de loi quelconque.

Étape n+1 :

$$\left\{ \begin{array}{l} \text{On tire } i_{n+1} \text{ dans la loi } Q(i_n, \cdot) = (q_{i_n j}, j \in E) \\ \text{On tire } u \text{ aléatoirement dans } [0, 1] \\ \text{Si } u < \alpha(i_n, i_{n+1}), \text{ on } \mathbf{accepte} \text{ la selection : } X_{n+1} = i_{n+1} \\ \text{Si } u > \alpha(i_n, i_{n+1}), \text{ on } \mathbf{refuse} \text{ la selection : } X_{n+1} = i_n \end{array} \right.$$

Remarque :

On peut facilement voir que l'algorithme de Metropolis initial est inclus dans l'algorithme de Metropolis généralisé en prenant Q symétrique et $h(u) = \min(1, u)$.

Théorème 2.3.1. *Supposons que la chaîne X de matrice de transition P et de loi initiale ν , produite par l'algorithme de Metropolis soit irréductible et ergodique. Alors :*

1. Pour tout $f \in l^1(\pi)$, on a :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{k=1}^N f(X_k) = \mathbb{E}_\pi(f) = \sum_{j \in E} f(j)\pi_j \quad \mathbb{P}_\nu\text{-p.s.}$$

2.

$$\mathcal{L}(X_n) \xrightarrow[n \rightarrow +\infty]{} \pi$$

Démonstration. Conséquence immédiate des théorèmes 1.6.2 et 1.6.3 : la chaîne étant ergodique, tous ses états sont récurrents positifs et aperiodiques. Ainsi 1 et 2 sont prouvés. \square

Ce théorème s'applique également aux cas où l'espace des états E est infini (en remplaçant la somme sur E par une intégrale). Il faut remplacer la condition d'irréductibilité par la condition de π -irréductibilité du noyau P (voir ci-dessous), qui sera alors vu comme une densité conditionnelle. On pourra trouver la justification de la convergence dans [ROBERT 96].

Définition 2.3.1. *Une chaîne de Markov X sur un espace d'état infini E est dite π -irréductible, si pour tout $A \in \mathcal{B}(X)$ tels que $\pi(A) > 0$, on a :*

$$\mathbb{P}_x(T_A < +\infty) > 0,$$

où

$$T_A = \inf\{n \geq 1, X_n \in A\}.$$

Théorème 2.3.2. [Peksun, 1973 [PEKSUN 73]]

Soient P_1 et P_2 deux matrices réversibles par rapport à π . On dit que P_2 est plus petit que P_1 ($P_2 \leq P_1$) si pour tout $x, y \in E$, tels que $x \neq y$, on a $P_2(x, y) \leq P_1(x, y)$.

Définissons également la variance asymptotique $v(f, \pi, P)$ de l'estimateur $S_N(f) = \frac{1}{N} \sum_{k=1}^N f(X_k)$ de $\mathbb{E}_\pi(f)$, construit à partir de f , et P de l'algorithme précédent par :

$$v(f, \pi, P) = \lim_{N \rightarrow +\infty} N \text{var} \left[\frac{1}{N} \sum_{k=1}^N f(X_k) \right].$$

Si $P_2 \leq P_1$, alors :

$$v(f, \pi, P_1) \leq v(f, \pi, P_2).$$

Proposition 2.3.2. La fonction d'acceptation de Metropolis et Hasting définie par $h^{MH}(u) = \min(1, u)$ est la plus grande parmi les fonctions $h :]0, +\infty[\rightarrow]0, 1[$ telles que $h(u) = uh(\frac{1}{u})$.

Par conséquent, d'après le théorème précédent, si P_1 est construit grâce au noyau de transition Q et la fonction h^{MH} , et P_2 grâce au même noyau de transition et à une autre fonction h , alors

$$v(f, \pi, P_1) \leq v(f, \pi, P_2).$$

Démonstration.

Si $u \geq 1$, alors $h^{MH}(u) = 1 \geq h(u)$ par définition de h .

Si $u \leq 1$, on a :

$$h(u) = uh\left(\frac{1}{u}\right) \leq u = h^{MH}(u).$$

□

2.4 Exemples

Nous présentons ici deux exemples d'utilisation des algorithmes de Métropolis. Les codes R correspondants seront placés en Annexe.

2.4.1 Simulation d'une loi de Poisson

Dans le but de simuler une loi de poisson grâce à l'algorithme de Metropolis, nous allons prendre pour Q la marche aléatoire sur \mathbb{N} , c'est à dire :

$$q_{ij} = \begin{cases} \frac{1}{2} & \text{si } i = j = 0 \\ \frac{1}{2} & \text{si } j = i + 1 \text{ ou } j = i - 1 \\ 0 & \text{sinon} \end{cases}$$

On peut voir qu'ainsi défini, Q est symétrique et irréductible.

$\pi_i \propto \frac{\lambda^i}{i!}$, par conséquent, quand $x=y=0$, le rapport $\frac{\pi_j}{\pi_i}$ vaut 1 et on accepte toujours y .

Si $i = 0$, $j = 1$ $\frac{\pi_j}{\pi_i} = \lambda > 1$ on accepte toujours j .

Si $j = i + 1$, $\frac{\pi_j}{\pi_i} = \frac{\lambda}{i}$

Si $j = i - 1$, $\frac{\pi_j}{\pi_i} = \frac{i}{\lambda}$.

L'algorithme s'écrit donc :

Etape 0 : Initialiser $X_0 = i$

Etape n+1 :

1. Si $i \neq 0$, on tire j aléatoirement entre $i - 1$ et $i + 1$. Sinon on tire j entre i et $i + 1$.
2. On tire u aléatoirement dans $[0, 1]$.
3. On calcule $\alpha = \frac{\pi_j}{\pi_i}$. Si $\alpha < u$, on accepte l'état j : $X_{n+1} = j$. Sinon $X_{n+1} = X_n$.

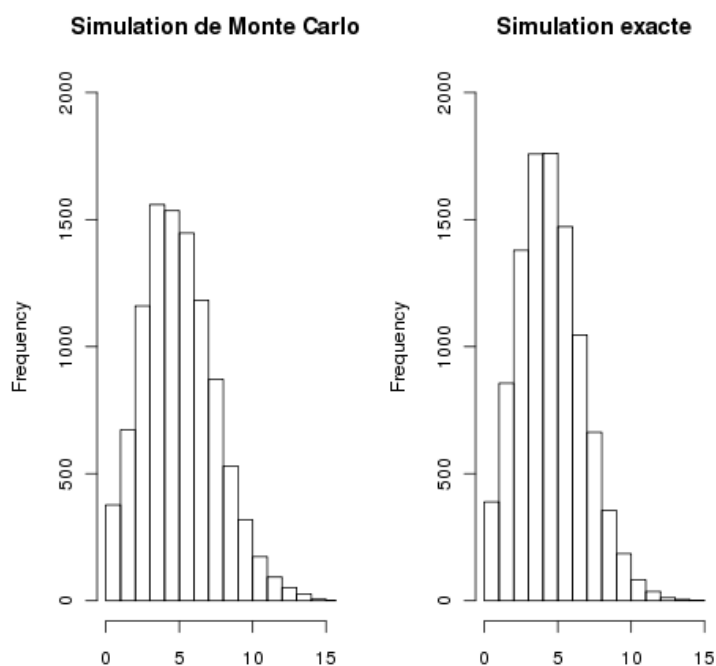


FIGURE 2.1 – Simulation d'une loi de poisson de paramètre $\lambda = 5$ avec 10000 itérations

Remarque :

Dans ce cas, $p_{00} = 1 - q_{01} = 1 - 1/2 = 1/2 \neq 0$ par conséquent, P est apériodique et récurrente positive, donc la convergence de la loi de X_n vers π est assurée pour n suffisamment grand d'après le théorème 1.6.3.

2.4.2 Modèle des sphères dures

Soit S un sous-ensemble de \mathbb{N} . Une configuration de *sphères dures* est une suite finie :

$$x = (x(s))_{s \in S}$$

où les $x(s)$ sont appelés *sites* et sont définis par :

$$x(s) \in \{0, 1\} \quad \forall s \in S \text{ et } x(u)x(v) = 0$$

pour tout sites u, v adjacents. Le diamètre d'une particule étant non-négligeable, deux particules ne peuvent occuper des positions voisines.

La condition $x(s) = 1$ (resp. $x(s) = 0$) s'interprète comme la présence (resp. l'absence) d'une particule au site s .

Soit l'espace des états $E \subset \{0, 1\}^{|S|}$, l'ensemble des configurations de sphères dures et π la probabilité uniforme sur E :

$$\forall i \in E, \pi_i = \frac{1}{|E|}$$

Cherchons une estimation du nombre moyen de sphères : $\mathbb{E}_\pi \left(\sum_{s \in S} x(s) \right)$

Pour cela nous avons recourt à l'algorithme de Metropolis :

Choisissons la fonction h définie par $h(u) = \frac{u}{1+u}$ et comme à chaque étape de l'itération, nous voulons qu'un seul site ne soit changé, nous prenons pour matrice de sélection, $Q : M * M \rightarrow [0, 1]$ définie par :

$$q_{ij} = \begin{cases} \frac{1}{|S|} & \text{si } x \text{ et } y \text{ diffèrent en un seul site} \\ 0 & \text{si } x \text{ et } y \text{ diffèrent en au moins deux sites} \end{cases}$$

Ainsi :

$$h \left(\frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right) = h(1) = \frac{1}{2}$$

L'algorithme mis en place est donc le suivant :

Etape 0 : Initialiser $X_0 = (0, \dots, 0)$

Etape n+1 :

1. On tire un sommet s au hasard parmi S . On retire s jusqu'à ce que tous ses voisins soient libres ($x(s+1) = x(s-1) = 0$).

2. On tire u aléatoirement dans $[0, 1]$

3.

$$\begin{cases} \text{Si } u < \frac{1}{2}, X_{n+1}(s) = 1 \\ \text{Si } u > \frac{1}{2}, X_{n+1}(s) = 0 \end{cases}$$

Q ainsi définie, tous les états peuvent être atteint à partir de n'importe quel état initial, par conséquent la chaîne ainsi produite est irréductible et apériodique (voir la proposition 1.4.1). Enfin en prenant la moyenne de la somme de chaque vecteur x obtenu par la chaîne de Markov, quand le nombre d'étape est suffisamment grande, on obtient alors une estimation du nombre moyen de sphère.

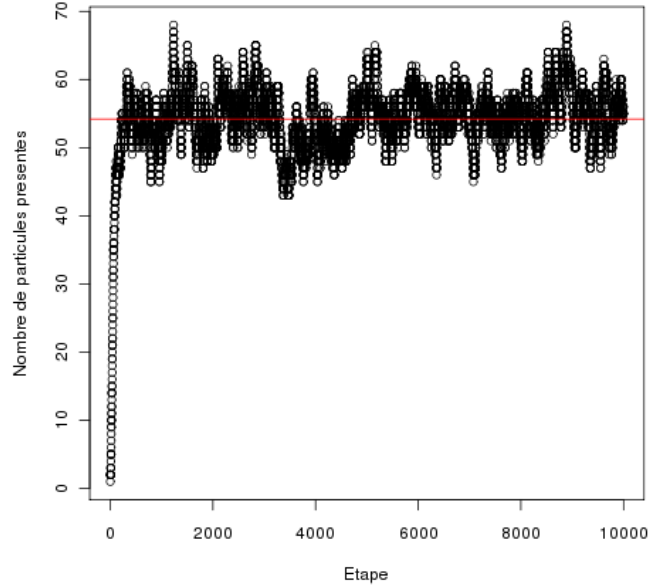


FIGURE 2.2 – Simulation d’un modèle de sphères dures par algorithme de Metropolis avec 10000 itérations pour une taille $S=200$.

En utilisant le code R placé en Annexe, on obtient un nombre moyen de sphères dures estimé à 54.5983.

Sur la figure 2.4.2, on peut légitimement se demander s’il peut être pertinent d’enlever la ”burn-in-period” pour une meilleure évaluation du nombre moyen de particules présentes. Dans ce cas, le choix de l’état initial influe très peu sur notre calcul puisque enlever les 500 premières étapes fait passer notre nombre moyen à 55,1 (au lieu de 54,6 avec).

2.5 Vitesse de convergence

Nous allons dans cette partie, regarder une technique d’étude de la vitesse de convergence des algorithmes de Metropolis, dans le cas de chaînes de Markov finies que l’on pourra retrouver dans [BENAÏM 05]. Ces techniques d’analyse spectrale permettent de déterminer le nombre de pas de simulation nécessaires à un algorithme de Metropolis, pour que la loi de la chaîne de Markov simulée soit proche de la loi π cherchée à un seuil ε donné. En effet, bien que les théorèmes ergodiques du chapitre 1 garantissent la convergence de ces algorithmes, une étude de la vitesse est nécessaire dans la mesure où rien ne garantit que la mesure invariante sera atteinte en un temps fini.

Soient E un espace d’états **fini** de cardinal d , π une probabilité sur E qui charge tous les états de E , et P un noyau d’une chaîne de Markov sur E ,

admettant π comme probabilité invariante.

Notons $l^2(\pi)$ l'espace vectoriel euclidien de dimension d , constitué des fonctions $f : E \rightarrow \mathbb{R}$ et muni du produit scalaire :

$$\langle f, g \rangle = \sum_{j \in E} f(j)g(j)\pi_j,$$

avec pour norme :

$$\|f\|_2 = \sqrt{\langle f, f \rangle}.$$

Nous noterons :

$$\pi f = \sum_{j \in E} f(j)\pi_j = \mathbb{E}_\pi(f).$$

Ainsi :

$$\text{var}_\pi(f) = \mathbb{E}_\pi((f - \pi f)^2) = \|f - \pi f\|_2^2.$$

Rappelons que nous avons :

$$Pf(i) = \sum_{j \in E} p_{ij}f(j).$$

Nous allons dans la section suivante, définir la notion de matrice adjointe qui nous sera utile pour la suite et établir quelques résultats concernant le spectre des matrices de transitions réversibles par rapport à loi π , qui seront utilisés pour le calcul de la vitesse de convergence des algorithmes.

2.5.1 Étude du spectre des matrices de transitions réversibles par rapport à π

Définition 2.5.1.

1. L'adjoint de P , noté P^* , est la matrice de transition définie par :

$$p_{ij}^* = \frac{\pi_j}{\pi_i} p_{ji}$$

Il vérifie :

$$\langle Pf, g \rangle = \langle f, P^*g \rangle \quad \forall f, g \in l^2(\pi)$$

Démonstration.

$$\begin{aligned} \langle Pf, g \rangle &= \sum_{i \in E} Pf(i)g(i)\pi_i \\ &= \sum_{i \in E} \sum_{j \in E} p_{ij}f(j)g(i)\pi_i \\ &= \sum_{j \in E} \pi_j f(j) \left(\sum_{i \in E} p_{ij}g(i) \frac{\pi_i}{\pi_j} \right) \\ &= \sum_{j \in E} \pi_j f(j) P^*g(j) = \langle f, P^*g \rangle \end{aligned}$$

□

2. P est dit **autoadjoint** si pour tout f et g dans $l^2(\pi)$, on a :

$$\langle Pf, g \rangle = \langle f, Pg \rangle.$$

Lemme 2.5.1. P est autoadjoint si et seulement si P est réversible par rapport à π

Démonstration. Supposons tout d'abord P réversible par rapport à π :

$$\pi_i p_{ij} = \pi_j p_{ji}.$$

Ainsi :

$$\begin{aligned} \langle Pf, g \rangle &= \sum_{i \in E} \sum_{j \in E} p_{ij} f(j) g(i) \pi_i \\ &= \sum_{j \in E} \sum_{i \in E} p_{ji} g(i) f(j) \pi_j = \langle f, Pg \rangle \end{aligned}$$

Ensuite si P est autoadjoint, pour tout f, g dans $l^2(\pi)$, on a $\langle Pf, g \rangle = \langle f, Pg \rangle$. Donc en prenant $f(i) = g(i) = 1$ pour tout $i \in E$, on obtient bien :

$$\pi_j p_{ji} = \pi_i p_{ij}.$$

□

Lemme 2.5.2. Si P est réversible par rapport à π , P est diagonalisable, et 1 est toujours valeur propre.

Démonstration. Si P est réversible, il est autoadjoint donc diagonalisable et ses valeurs propres sont réelles. De plus comme P est stochastique,

$$P \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Donc 1 est valeur propre, de vecteur propre le vecteur composé de 1. □

Proposition 2.5.1. Supposons que P soit réversible par rapport à π . Alors :

1. Les valeurs propres de P sont de module inférieur ou égal à 1. On peut donc ordonner ces valeurs propres :

$$1 = \beta_1 \geq \beta_2 \geq \dots \geq \beta_d \geq -1.$$

2. Si P est irréductible, 1 est valeur propre simple :

$$1 = \beta_1 > \beta_2 \geq \dots \geq \beta_d \geq -1.$$

3. Si P est irréductible et apériodique, -1 n'est pas valeur propre :

$$1 = \beta_1 > \beta_2 \geq \dots \geq \beta_d > -1.$$

Démonstration.

1. Soit f une fonction propre associée à la valeur propre β ($Pf = \beta f$). On a :

$$|\beta| \|f\|_\infty = \|Pf\|_\infty = \sup_{i \in E} |Pf(i)| \leq \sup_{i \in E} \sum_{j \in E} p_{ij} |f(j)| \leq \|f\|_\infty .$$

D'où $|\beta| \leq 1$

2. Supposons P irréductible et que 1 est valeur propre. Il suffit de montrer que P irréductible et $Pf(i) = f(i) \forall i \in E \Rightarrow f$ est constante et donc 1 est valeur propre simple. Pour cela nous utiliserons la notion de forme de Dirichlet que l'on retrouvera dans la partie suivante. Le corollaire 2.5.2 démontrera alors le résultat.
3. Supposons que -1 est valeur propre. On a alors $Pf(x) = -f(x)$ pour tout x dans E . D'où pour tout entier n , $P^{2n+1}f(x) = -f(x)$. P étant irréductible et apériodique, d'après le théorème ergodique 1.6.3 on a :

$$\lim_{n \rightarrow +\infty} P^{2n+1}f(x) = \lim_{n \rightarrow +\infty} \sum_{y \in E} p^{2n+1}(x, y) f(y) = \sum_{y \in E} \pi_j f(y) = -f(x) .$$

Ce qui est absurde. □

2.5.2 Vitesse de convergence des noyaux réversibles

Supposons pour la suite P réversible donc autoadjoint. P est donc diagonalisable dans une base orthonormée de fonctions propres : ψ_1, \dots, ψ_d où $\psi_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

Pour toute fonction $f \in l^2(\pi)$:

$$f = \sum_{i=1}^d \langle f, \psi_i \rangle \psi_i = (\pi f) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \sum_{i=2}^d \langle f, \psi_i \rangle \psi_i .$$

D'où pour tout $i \in E$, X_0 de loi π :

$$\mathbb{E}(f(X_n) | X_0 = i) - \mathbb{E}_\pi(f) = P^n f(i) - \pi f = \sum_{i=2}^d \beta_i^n \langle f(i), \psi_i \rangle \psi_i .$$

Théorème 2.5.1. *Soit P la matrice de transition d'une chaîne de Markov X sur E , réversible par rapport à π , irréductible et apériodique.*

Alors, pour toute fonction f dans $l^2(\pi)$ et pour tout état initial i :

$$(\mathbb{E}_i(f(X_n)) - \mathbb{E}_\pi(f))^2 = (P^n f(i) - \pi f)^2 \leq c(i) \rho^{2n+2} \text{var}_\pi(f),$$

où

$$c(i) = \frac{p_{ii}^2}{\pi_i} \quad \rho = \rho(P) = \sup(\beta_2, |\beta_d|) < 1 \quad (\text{d'après 2.5.1(3)}).$$

Démonstration. Prenons $g = f - \mathbb{E}_\pi(f)$. Ainsi $\pi g = 0$. D'une part :

$$\|P^n g\|_2^2 = \sum_{i=2}^d \beta_i^{2n} \langle g, \psi_i \rangle^2 \underbrace{\|\psi_i\|_2^2}_{=1} \leq \rho^{2n} \|g\|_2^2 = \rho^{2n} \text{var}_\pi(g).$$

Ensuite Posons $\Phi = P^{n-1}g$. Ainsi comme π est réversible :

$$P\Phi(i) = \sum_{j \in E} p_{ij} \Phi(j) = \sum_{j \in E} \frac{p_{ji}}{\pi_i} \Phi(j) \pi_j = \langle g_i, \Phi \rangle,$$

où g_i est définie par $g_i(j) = \frac{p_{ji}}{\pi_i}$

en utilisant l'inégalité de Cauchy-Schwarz, on obtient :

$$|P\Phi(i)|^2 = |\langle g_i, \Phi \rangle|^2 \leq \|g_i\|_2^2 \|\Phi\|_2^2.$$

Or :

$$\begin{aligned} \|g_i\|_2^2 &= \sum_{j \in E} \left(\frac{p_{ji}}{\pi_i} \right)^2 \pi_j \\ &= \sum_{j \in E} \frac{p_{ji}}{\pi_i} p_{ji} \pi_j \\ &= \frac{1}{\pi_i} \sum_{j \in E} p_{ji} p_{ij} \text{ par réversibilité} \\ &= \frac{p_{ii}^2}{\pi_i} \text{ d'après les équations de Chapman-Kolmogorov.} \end{aligned}$$

Par conséquent,

$$|P\Phi(i)|^2 \leq c(i) \|\Phi\|_2^2 \leq c(i) \rho^{2n-2} \text{var}_\pi(g).$$

□

Corollaire 2.5.1. *Sous les mêmes hypothèses, pour tout état initial $i \in E$ et pour tout $A \subset E$:*

$$|\mathbb{P}(X_n \in A) - \pi(A)|^2 \leq \frac{1}{4} c(i) \rho^{2n+2}.$$

Démonstration. Il suffit d'appliquer le théorème précédent à la fonction indicatrice sur A , en remarquant que :

$$\text{var}_\pi(1_A) = \pi(A)(1 - \pi(A)) \leq \frac{1}{4}.$$

□

Ce corollaire nous permet de calculer le nombre de pas nécessaires pour avoir la précision voulue avec l'algorithme de Métropolis-Hasting.

Cependant le calcul de ρ est, dans la plupart des cas, impossible à cause de la taille trop importante de P . C'est pourquoi nous allons voir un autre moyen de calculer ce nombre de pas en utilisant la notion de trou spectral.

2.5.3 Formes de Dirichlet et trou spectral

Définition 2.5.2. La *forme de Dirichlet* d'un noyau P d'une chaîne X est définie pour tout $f \in l^2(\pi)$ par :

$$\varepsilon(f, f) = \langle (I - P)f, f \rangle .$$

Lemme 2.5.3. Pour tout $f \in l^2(\pi)$:

$$\varepsilon(f, f) = \langle (I - P)f, f \rangle = \langle (I - \frac{P + P^*}{2})f, f \rangle = \frac{1}{2} \sum_{i \in E} \sum_{j \in E} (f(i) - f(j))^2 p_{ij} \pi_i .$$

Démonstration. Remarquons que la première égalité est immédiate, puisque P^* est l'adjoint de P .

Pour prouver la seconde, soit X_0 de loi π . On a :

$$\begin{aligned} \frac{1}{2} \sum_{i \in E} \sum_{j \in E} (f(i) - f(j))^2 p_{ij} \pi_i &= \frac{1}{2} \sum_{i \in E} \sum_{j \in E} (f(i) - f(j))^2 \mathbb{P}(X_1 = j, X_0 = i) \\ &= \frac{1}{2} \mathbb{E}((f(X_0) - f(X_1))^2) \\ \text{Comme } \pi \text{ est invariant} &= \mathbb{E}(f(X_0)^2) - \mathbb{E}((f(X_1)f(X_0))) \\ &= \mathbb{E}(f(X_0)^2) - \mathbb{E}(\mathbb{E}(f(X_1)|X_0)f(X_0)) \\ &= \mathbb{E}(f(X_0)^2) - \mathbb{E}(Pf(X_0)f(X_0)) \\ &= \mathbb{E}((f(X_0) - Pf(X_0))f(X_0)) = \varepsilon(f, f) . \end{aligned}$$

□

Corollaire 2.5.2. Si P est irréductible, les fonctions f P -invariantes ($Pf = f$) sont les fonctions constantes.

Démonstration. Comme P est irréductible et π charge tous les points de E , d'après la dernière égalité du lemme 2.5.3, toute fonction annulant la forme de Dirichlet est nécessairement constante. Or de part la définition de la forme de Dirichlet, il est clair que si f est P -invariante, alors $\varepsilon(f, f) = 0$ d'où f est constante. □

Définition 2.5.3. Le *trou spectral* du noyau P est défini par :

$$\lambda = \lambda(P) := \inf_f \left\{ \frac{\varepsilon(f, f)}{\text{var}_\pi(f)} ; \text{var}_\pi(f) \neq 0 \right\} .$$

La forme de Dirichlet étant une forme positive d'après le lemme 2.5.3, $\lambda \geq 0$. Remarquons que λ peut s'écrire :

$$\lambda = \inf_f \{ \varepsilon(f, f), \pi f = 0, \|f\|_2 = 1 \} .$$

Donc en diagonalisant $I - \frac{P + P^*}{2}$ dans une base orthonormale, on peut conclure

que λ est la plus petite valeur propre non-nulle de $I - \frac{P + P^*}{2}$. Par conséquent, si π est réversible, λ est la plus petite valeur propre non-nulle de $I - P$, i.e : $\lambda = 1 - \beta_2$.

Proposition 2.5.2. *Si P est réversible et $p_{ii} \geq \frac{1}{2}$ pour tout i appartenant à E , alors toutes les valeurs propres de P sont positives. (et donc $\rho = \beta_2 = 1 - \lambda$)*

Démonstration. Soient β une valeur propre de P de vecteur propre f , et i tel que $|f(i)| \geq |f(j)|$ pour tout j dans E . Alors :

$$\sum_{j \neq i} p_{ij} f(j) = Pf - p_{ii} f(i) = (\beta - p_{ii}) f(i),$$

ainsi :

$$|\beta - p_{ii}| \leq \sum_{j \neq i} p_{ij} \frac{|f(j)|}{|f(i)|} \leq 1 - p_{ii}.$$

Par conséquent

$$\beta - p_{ii} \geq p_{ii} - 1 \Rightarrow \beta \geq 2p_{ii} - 1 \geq 0.$$

□

Proposition 2.5.3. *Si P est réversible et positif ($\langle Pf, f \rangle \geq 0$ pour tout f), alors ici encore toutes les valeurs propres de P sont positives.*

Démonstration. Soit β une valeur propre de P de vecteur propre f .

$$\langle Pf, f \rangle = \langle \beta f, f \rangle = \beta \langle f, f \rangle = \beta \|f\|_2^2 \geq 0$$

Comme $\|f\|_2^2 \geq 0$, $\beta \geq 0$.

□

Introduisons à présent la norme $\|\cdot\|_p$ dans $l^p(\pi)$ définie par :

$$\|h\|_p := \left(\sum_{j \in E} |h(j)|^p \pi_j \right)^{\frac{1}{p}}.$$

On définit alors la distance du χ^2 entre deux mesures μ et π lorsque $\pi > 0$ par :

$$\chi(\mu, \pi) := \left\| \frac{\mu}{\pi} - 1 \right\|_2.$$

Lemme 2.5.4. *Soit $|\mu - \pi|_{VT} = \frac{1}{2} \sum_{j \in E} |\pi_j - \mu_j|$ la **distance en variation totale** entre μ et π . On a :*

$$\frac{1}{2} \left\| \frac{\mu}{\pi} - 1 \right\|_1 = |\mu - \pi|_{VT}.$$

Soit $h(x)$ une fonction dans $l^2(\pi)$. En utilisant l'inégalité de Jensen concave, On a :

$$\|h - 1\|_1 = \mathbb{E}(|h - 1|) \leq \mathbb{E}^{\frac{1}{2}}(|h - 1|^2) = \|h - 1\|_2.$$

Par conséquent en prenant $h = \frac{\mu}{\pi}$:

$$|\mu - \pi|_{VT} \leq \frac{1}{2} \chi(\mu, \pi).$$

Donc majorer la distance du χ^2 entre la loi de X_n et π nous donnera une bonne estimation du nombre de pas nécessaires pour obtenir la précision voulue. C'est ce que nous donne le théorème suivant :

Théorème 2.5.2. *Soit P une matrice de transition d'une chaîne de Markov sur un espace état E fini, telle que P soit apériodique, irréductible et de loi invariante π (pas nécessairement réversible). Alors pour toute loi μ sur E , on a :*

$$\chi^2(\mu P^n, \pi) \leq (1 - \lambda(PP^*))^n \chi^2(\mu, \pi).$$

Démonstration. Notons ε' la forme de Dirichlet associée au noyau PP^* , et $\rho_n(i) = \frac{\mu P^n(i)}{\pi_i}$. Remarquons tout d'abord que $P^* \rho_n = \rho_{n+1}$:

On a :

$$\rho_n(i) = \sum_{j \in E} \frac{\mu_j p_{ij}(n)}{\pi_i},$$

et

$$\begin{aligned} P^* \rho_n &= \sum_{j \in E} p_{ij}^* \rho_n(j) \\ &= \sum_{j \in E} p_{ij}^* \sum_{z \in E} \frac{\mu_z p_{jz}(n)}{\pi_j} \\ &= \sum_{j \in E} \frac{\pi_j}{\pi_i} p_{ij} \sum_{z \in E} \frac{\mu_z p_{jz}(n)}{\pi_j} \\ &= \sum_{z \in E} \frac{\mu_z}{\pi_i} \sum_{j \in E} p_{ij} p_{jz}(n) \\ &= \sum_{z \in E} \frac{\mu_z}{\pi_i} p_{iz}(n+1) \text{ d'après les équations de Chapman-Kolmogorov} \\ &= \rho_{n+1}. \end{aligned}$$

D'où par invariance de π , on a $\mathbb{E}_\pi(\rho_n) = \mathbb{E}_\pi(P^* \rho_n) = 1$.

Ainsi puisque P et P^* sont adjoints :

$$\varepsilon'(\rho_n) = \langle \rho_n, (I - PP^*) \rho_n \rangle = \langle \rho_n, \rho_n \rangle - \langle P^* \rho_n, P^* \rho_n \rangle.$$

D'où :

$$\varepsilon'(\rho_n) = \text{var}_\pi(\rho_n) - \text{var}_\pi(P^* \rho_n) = \text{var}_\pi(\rho_n) - \text{var}_\pi(\rho_{n+1}).$$

Et donc par définition du trou spectral :

$$\text{var}_\pi(\rho_n) = \varepsilon'(\rho_n) + \text{var}_\pi(\rho_{n+1}) \geq \lambda(PP^*) \text{var}_\pi(\rho_n) + \text{var}_\pi(\rho_{n+1}).$$

En utilisant la définition de χ , on a également que :

$$\chi^2(\mu P^n, \pi) = \|\rho_n - 1\|_2^2 = \text{var}_\pi(\rho_n),$$

ainsi :

$$\chi^2(\mu P^n, \pi) \leq (1 - \lambda(PP^*)) \chi^2(\mu P^n, \pi).$$

□

Ainsi, si on simule un algorithme de Metropolis-Hasting en initialisant X_0 de loi μ , la distance en terme de variations totales entre la loi de X_n et π sera

inférieur à $\frac{1}{2}(1 - \lambda(P P^*))^{\frac{\alpha}{2}} \chi(\mu, \pi)$.

Nous retrouverons dans [BENAIM 05] des exemples d'utilisations de ces méthodes spectrales. Bien que nous exposons ces méthodes spectrales, nous ne les utiliserons pas dans la suite de ce mémoire, cependant à terme, elles seront très certainement utilisées afin de calculer de façon précise, la vitesse de convergence de nos algorithmes pour la simulation de la lumière dans les tissus du cerveau.

Nous allons à présent aborder le problème de la diffusion de la lumière, et nous allons voir de quel façon ce problème pourra être résolu en utilisant une adaptation des algorithmes de Métropolis décrits dans ce chapitre.

Chapitre 3

Metropolis pour la synthèse d'image

Nous allons dans ce chapitre étudier les algorithmes MLT ("Metropolis Light Transport") décrits par Veach [VEACH 97] et repris par Pauly [PAULY 00], initialement développés dans le domaine de la synthèse d'images. Nous nous en servons comme alternative à la méthode de Monte Carlo simulant la propagation de photons dans les tissus biologiques développée dans [WANG 95]. La méthode MLT offre une meilleure perspective en terme de temps de calcul et de parallélisation.

3.1 Metropolis Light Transport

Le problème du transport de la lumière sur un domaine $D \in \mathbb{R}^3$ est généralement vu comme une équation intégrale, où nous devons chercher à trouver la fonction de radiance L . Nous supposons dans cette partie que la lumière est émise et absorbée seulement aux surfaces, se déplaçant en ligne droite entre ces surfaces.

3.1.1 Equations de transport de la lumière

Soit \mathcal{M} l'ensemble des points de notre scène d'étude pour lesquels la lumière peut être émise ou absorbée (l'union des surfaces d'étude). Dans ce cas, la radiance au point x_1 dans la direction x_2 est définie par :

$$\begin{aligned} L : \mathcal{M} \times \mathcal{M} &\rightarrow \mathbb{R}_+ \\ (x_1, x_2) &\mapsto L_e(x_1, x_2) + \int_{\mathcal{M}} L(x, x_1) f_s(x, x_1, x_2) G(x, x_1) dA(x), \end{aligned}$$

où A désigne la surface de mesure sur \mathcal{M} , et L_e désigne la radiance émise en partant de x_1 dans la direction x_2 .

La fonction G est définie par :

$$G(x, y) = V(x, y) \frac{\cos(\theta_x) \cos(\theta_y)}{\|x - y\|^2}, \quad (3.1)$$

θ_x et θ_y désignant les angles entre le segment $[xy]$ et la surface normale à x et y respectivement, tandis que $V(x, y)$ vaut 1 si x et y sont mutuellement visibles and 0 sinon. La fonction f_s est la fonction de distribution de diffusion bidirectionnelle (BDSF, bidirectional scattering distribution function). Cette fonction décrit les directions possibles que peut prendre la lumière après avoir frappé une surface, elle dépend donc du domaine d'étude et sera détaillée dans notre cas dans le chapitre suivant. Elle prendra en argument trois points x_1, x_2, x_3 de \mathcal{M} ou plus exactement l'angle formé par les directions $\overrightarrow{x_1x_2}$ et $\overrightarrow{x_2x_3}$.

Il y a une redondance dans la représentation de L ; si x_1 et x_2 sont dans la même direction en partant de x , alors $L(x, x_1) = L(x, x_2)$.

Nous pouvons voir que L peut être réécrit grâce à un opérateur intégral T sur \mathcal{M} :

$$L(x_1, x_2) = L_e(x_1, x_2) + TL(x_1, x_2),$$

où

$$TL(x_1, x_2) = \int_{\mathcal{M}} L(x, x_1) f_s(x, x_1, x_2) G(x, x_1) dA(x).$$

Ainsi

$$L(x_1, x_2) = (I - T)^{-1} L_e(x_1, x_2).$$

Si l'on suppose : $\|T\| < 1$, ([VEACH 97])

$$L(x_1, x_2) = \sum_{j=0}^{+\infty} T^j L_e(x_1, x_2), \quad (3.2)$$

où chaque T^j est un opérateur intégral sur \mathcal{M}^j .

3.1.2 Estimation de l'intensité de lumière m_j

Plutôt que de chercher à estimer la fonction de radiance L , Metropolis Light Transport est un algorithme visant à estimer la valeur de chaque pixel m_1, \dots, m_M , d'une image. Chacune des valeurs de ces pixels à la forme ([VEACH 97]) :

$$m_j = \int_{\mathcal{M} \times \mathcal{M}} W_e^{(j)}(x, x_0) L(x, x_0) G(x, x_0) dA(x) dA(x_0) \quad (3.3)$$

Où $W_e^{(j)}(x, x_0)$ est un poids indiquant la façon dont la lumière arrivant au point x_0 en partance du point x contribue à la valeur de la mesure. La fonction $W_e^{(j)}(x, y)$ vaut zéro presque partout, et est non-nulle si et seulement si y est situé dans une petite région de l'image correspondant au support du filtre du pixel j .

En remplaçant L dans l'équation (3.3) grâce à l'équation (3.2), on obtient :

$$\begin{aligned} m_j &= \int_{\mathcal{M}^2} W_e^{(j)}(x_1, x_0) L_e(x_1, x_0) G(x_1, x_0) dA(x_1) dA(x_0) \\ &+ \int_{\mathcal{M}^3} W_e^{(j)}(x_1, x_0) L_e(x_2, x_1) G(x_1, x_0) f_s(x_2, x_1, x_0) G(x_2, x_1) dA(x_2) dA(x_1) dA(x_0) \\ &+ \dots \end{aligned}$$

Ainsi :

$$m_j = \sum_{k=1}^{+\infty} \left[\int_{\mathcal{M}^{k+1}} W_e^{(j)}(x_1, x_0) \left(\prod_{l=1}^{k-1} G(x_l, x_{l-1}) f_s(x_{l+1}, x_l, x_{l-1}) \right) L_e(x_k, x_{k-1}) G(x_k, x_{k-1}) dA(x_k) \dots dA(x_0) \right].$$

On peut réécrire m_j de la façon suivante, où on a échangé les rôles de x_0 et de x_k , de x_1 et de x_{k-1} , etc. en se rappelant que les x_0, \dots, x_k ne sont rien d'autre que des variables d'intégration :

$$\begin{aligned} m_j &= \sum_{k=1}^{+\infty} \left[\int_{\mathcal{M}^{k+1}} W_e^{(j)}(x_{k-1}, x_k) \left(\prod_{l=1}^{k-1} G(x_{k-l}, x_{k-l+1}) f_s(x_{k-l-1}, x_{k-l}, x_{k-l+1}) \right) L_e(x_0, x_1) G(x_0, x_1) dA(x_0) \dots dA(x_k) \right] \\ &= \sum_{k=1}^{+\infty} \left[\int_{\mathcal{M}^{k+1}} W_e^{(j)}(x_{k-1}, x_k) \left(\prod_{l=1}^{k-1} G(x_l, x_{l+1}) f_s(x_{l-1}, x_l, x_{l+1}) \right) L_e(x_0, x_1) G(x_0, x_1) dA(x_0) \dots dA(x_k) \right] \end{aligned}$$

Le but est de réécrire l'expression m_j sous la forme d'une seule intégrale sur un espace Ω par rapport à une mesure complexe μ , dans l'idée d'utiliser un algorithme de Métropolis que l'on détaillera plus après :

$$\boxed{m_j = \int_{\Omega} f_j(\bar{x}) d\mu(\bar{x})} \quad (3.4)$$

Avec \bar{x} une trajectoire appartenant à Ω . Pour faire ainsi, nous définissons : Ω_k l'ensemble de toutes les trajectoires de tailles k (la taille d'une trajectoire fera référence au nombre d'arrête et non au nombre de sommet) : $\Omega_k = \mathcal{M}^{k+1}$, et les mesures μ_k sur Ω_k par :

$$d\mu_k(x_0, \dots, x_k) = dA(x_0) \dots dA(x_k)$$

Nous prenons alors Ω l'ensemble de toutes les trajectoires : $\Omega = \cup_{k \geq 1} \mathcal{M}^{k+1}$, et définissons une mesure μ sur Ω par :

$$\mu(D) = \sum_{k=1}^{+\infty} \mu_k(D \cap \Omega_k) = \sum_{k=1}^{+\infty} A^{\otimes k}(D \cap \Omega_k) \quad \text{pour } D \subset \Omega$$

Nous définissons alors les fonctions f_j^k allant de Ω_k dans \mathbb{R}^+ par

$$f_j^k(x_0 x_1 \dots x_k) = L_e(x_0, x_1) G(x_0, x_1) \left(\prod_{l=1}^{k-1} G(x_l, x_{l+1}) f_s(x_{l-1}, x_l, x_{l+1}) \right) W_e^{(j)}(x_{k-1}, x_k),$$

où $x_0x_1\dots x_k$ désigne le chemin(ou trajectoire) lumineux allant du point x_0 au point x_k . Ainsi, f_j sera une fonction de Ω dans \mathbb{R}^+ et sera définie par :

$$f_j(\bar{x}) = \sum_{l=1}^{+\infty} f_j^l(x_0 \dots x_l) \mathbf{1}_{(\bar{x} \in \mathcal{M}^{l+1})}$$

Par exemple, on a :

$$\begin{aligned} f_j(x_0x_1) &= L_e(x_0, x_1)G(x_0, x_1)W_e^{(j)}(x_0, x_1) \\ f_j(x_0x_1x_2) &= L_e(x_0, x_1)G(x_0, x_1)f_s(x_0, x_1, x_2)G(x_1, x_2)W_e^{(j)}(x_1, x_2) \\ f_j(x_0x_1x_2x_3) &= L_e(x_0, x_1)G(x_0, x_1)f_s(x_0, x_1, x_2)G(x_1, x_2)f_s(x_1, x_2, x_3)W_e^{(j)}(x_2, x_3) \end{aligned}$$

On obtient bien :

$$\int_{\Omega} f_j(\bar{x}) d\mu(\bar{x}) = \sum_{k=1}^{+\infty} \left[\int_{\mathcal{M}^{k+1}} f_j^k(x_0 \dots x_k) d\mu_k(\bar{x}) \right] = m_j$$

où on a permutté somme et intégrale par Fubini.

Nous pouvons aisément voir qu'on peut réécrire f_j de la façon suivante :

$$f_j(\bar{x}) = w_j(\bar{x})f(\bar{x}) \tag{3.5}$$

$$\begin{aligned} w_j(\bar{x}) &= \sum_{k=1}^{+\infty} W_e^{(j)}(x_{k-1}, x_k) \mathbf{1}_{(\bar{x} \in \mathcal{M}^{k+1})} \\ f(\bar{x}) &= \sum_{k=1}^{+\infty} L_e(x_0, x_1)G(x_0, x_1) \left(\prod_{l=1}^{k-1} G(x_l, x_{l+1})f_s(x_{l-1}, x_l, x_{l+1}) \right) \mathbf{1}_{(\bar{x} \in \mathcal{M}^{k+1})} \end{aligned}$$

3.2 Influence du milieu

Dans cette partie, on suppose qu'entre les surfaces, la lumière peut être absorbée et diffusée dans le milieu même dans laquelle elle circule (par exemple si le milieu est constitué de brouillard, fumée ou feu). Dans notre cas, la lumière circulera dans les tissus du cerveau, et on ne pourra pas négliger l'impact du milieu sur la réfraction de la lumière.

Nous allons devoir considérer le cas où la lumière est réfléchiée par les surfaces, et le cas où la lumière est réfléchiée par un milieu. Chaque milieu est décrit par un coefficient d'absorption σ_a et un coefficient de diffusion σ_s , qui peuvent être variables en fonction de la position dans le milieu (milieu inhomogène). Nous allons supposer être dans le cas d'un volume fini $\mathcal{V} \subset \mathbb{R}^3$. Nous noterons $\partial\mathcal{V}$ la frontière de \mathcal{V} , c'est à dire, l'ensemble fini des surfaces décrivant les objets de l'espace de notre étude. L'espace entre les objets constituera donc $\mathcal{V}^0 = \mathcal{V} \setminus \partial\mathcal{V}$

Nous allons définir $\pi_{-\omega}(x)$ le point appartenant à $\partial\mathcal{V}$, issu de $x \in \mathcal{V}^0$ dans une direction $-\omega \in \mathcal{S}^2$, \mathcal{S}^2 étant la sphère unité. Nous définissons également un coefficient d'extinction défini par $\sigma_e = \sigma_a + \sigma_s$ et la fonction $\tau(x, y)$ (path transmittance) entre $x, y \in \mathcal{V}^0$ définie par :

$$\tau(x, y) = \exp\left(-\int_x^y \sigma_e(z) dz\right).$$

Notons aussi $L_{e, \partial\mathcal{V}}$ la radiance émise par les surfaces, et L_{e, \mathcal{V}^0} la radiance émise dans le volume.

Pauly montre dans [PAULY 00] que tenant compte des conditions sur la frontière, la fonction de radiance L peut être écrite :

$$L = (L_{i, \partial\mathcal{V}} + L_{i, \mathcal{V}^0}) + (T_{\partial\mathcal{V}} + T_{\mathcal{V}^0})L,$$

où :

$$\begin{aligned} L_{i, \partial\mathcal{V}}(x, \omega) &= \tau(\pi_{-\omega}(x), x) L_{e, \partial\mathcal{V}}(\pi_{-\omega}(x), x) \\ L_{i, \mathcal{V}^0}(x, \omega) &= \int_{\pi_{-\omega}(x)}^x \tau(z, x) L_{e, \mathcal{V}^0}(z, \omega) dz \end{aligned}$$

et les opérateurs $T_{\partial\mathcal{V}}$ et $T_{\mathcal{V}^0}$ sont définis par :

$$\begin{aligned} T_{\partial\mathcal{V}}L(x, \omega) &= \tau(\pi_{-\omega}(x), x) \int_{\mathcal{S}^2} f_s(\omega, \pi_{-\omega}(x), \omega') L(\pi_{-\omega}(x), \omega') \cos(\Theta_{\pi_{-\omega}(x)}) d\omega' \\ T_{\mathcal{V}^0}L(x, \omega) &= \int_{\pi_{-\omega}(x)}^x \tau(z, x) \sigma_s(z) \int_{\mathcal{S}^2} f_s(\omega, z, \omega') L(z, \omega') d\omega' dz, \end{aligned}$$

f_p étant la fonction de phase, qui décrit les caractéristiques de diffusion du milieu, et f_s la fonction BDSF décrite en partie dans la partie précédente et qui sera détaillée dans le chapitre suivant.

En supposant que $\|T_{\partial\mathcal{V}} + T_{\mathcal{V}^0}\| < 1$ (dans [PAULY 00], Pauly montre que c'est le cas dans la plupart des cas), de manière analogue à la première partie on peut réécrire L de la façon suivante :

$$\begin{aligned}
L &= \sum_{j=0}^{+\infty} (T_{\partial\mathcal{V}} + T_{\mathcal{V}^0})^j (L_{i,\partial\mathcal{V}} + L_{i,\mathcal{V}^0}) \\
&= L_{i,\partial\mathcal{V}} + L_{i,\mathcal{V}^0} + T_{\partial\mathcal{V}}L_{i,\partial\mathcal{V}} + T_{\mathcal{V}^0}L_{i,\mathcal{V}^0} + \dots \quad (3.6)
\end{aligned}$$

De la même façon que dans la partie précédente, nous allons chercher à calculer la valeur de chaque pixel m_j , $j = 1, \dots, M$ d'une image de taille M . Pour cela, nous pouvons exprimer m_j par :

$$\begin{aligned}
m_j &= \int_{\mathcal{S}^2 \times \mathcal{V}} W_e^{(j)}(x, \omega) L(x, \omega) d\omega dV(x) \\
&= \int_{\mathcal{V} \times \mathcal{V}} W_e^{(j)}(x_0, x_1) L(x_0, x_1) G(x_0, x_1) dV(x_0) dV(x_1),
\end{aligned}$$

avec :

$$G(x, y) = V(x, y) \frac{D_x(y) D_y(x)}{\|y - x\|^2} \tau(x, y),$$

où $D_x(y)$ vaut 1 si $x \in \mathcal{V}^0$ et sinon vaut la valeur absolue du produit scalaire entre la direction ω_{xy} et la surface normale à x . De même pour $D_y(x)$.

De la même façon que dans la partie 1, nous allons injecter l'équation (3.6) dans l'expression précédente des m_j et nous allons à nouveau exprimer m_j comme l'intégrale sur Ω d'une fonction f_j par rapport à une mesure μ .

Pour se faire nous allons poser Ω_k^l l'ensemble des chemins de longueur k , où $l = (l_0, \dots, l_k)$, tels que pour $j = 0, \dots, k$, $x_j \in \partial\mathcal{V}$ si $l_j = 1$ et $x_j \in \mathcal{V}^0$ si $l_j = 0$.

Par exemple : $\Omega_3^{(1101)}$ est l'ensemble des chemins de longueur 3 tels que x_0, x_1 et x_3 appartiennent aux frontières et x_2 est dans le volume.

On choisira alors :

$$\Omega = \bigcup_{k=1}^{+\infty} \bigcup_{l \in \{0,1\}^{k+1}} \Omega_k^l.$$

Une mesure μ_k^l sur Ω_k^l sera donnée par :

$$d\mu_k^l(x_0, \dots, x_k) = d\mu_0^l(x_0) \dots d\mu_k^l(x_k)$$

où

$$d\mu_i^l(x_i) = \begin{cases} dA(x_i) & \text{si } l_i = 1 \\ dV(x_i) & \text{si } l_i = 0 \end{cases}.$$

Nous pouvons à présent définir une mesure μ sur l'ensemble des chemins Ω par :

$$\mu(D) = \sum_{k=1}^{+\infty} \sum_{l \in \{0,1\}^{k+1}} \mu_k(D \cap \Omega_k^l) \quad \text{pour } D \subset \Omega.$$

Ainsi de la même façon que précédemment,

$$f_j(\bar{x}) = \sum_{k=1}^{+\infty} \sum_{l \in \{0,1\}^{k+1}} f_j^k(x_0 \dots x_k) \mathbf{1}(\bar{x} \in \Omega_k^l),$$

avec pour tout entier k :

$$f_j^k(x_0, \dots, x_k) = L_e(x_0, x_1) G(x_0, x_1) \left(\prod_{l=1}^{k-1} G(x_l, x_{l+1}) \hat{f}(x_{l-1}, x_l, x_{l+1}) \right) W_e^{(j)}(x_{k-1}, x_k)$$

et pour tout $l = 1, \dots, k-1$:

$$\hat{f}(x_{l-1}, x_l, x_{l+1}) = \begin{cases} f_s(x_{l-1}, x_l, x_{l+1}) & \text{Si } x_l \in \partial\mathcal{V} \\ \sigma_s(x_l) f_p(x_{l-1}, x_l, x_{l+1}) & \text{Si } x_l \in \mathcal{V}^0 \end{cases},$$

ainsi que :

$$L_e(x_0, x_1) = \begin{cases} L_{e, \partial\mathcal{V}}(x_0, x_1) & \text{Si } x_0 \in \partial\mathcal{V} \\ L_{e, \mathcal{V}^0}(x_0, x_1) & \text{Si } x_0 \in \mathcal{V}^0 \end{cases}.$$

3.3 Algorithme de Metropolis

Nous allons voir comment l'algorithme de Metropolis peut être adapté pour estimer toutes les mesures m_j simultanément et sans biais.

À partir de (3.4) et (3.5) nous avons :

$$m_j = \int_{\Omega} w_j(\bar{x}) f(\bar{x}) d\mu(\bar{x}) \quad (3.7)$$

f étant une fonction allant de Ω dans \mathbb{R}^+ , $f(x)\mu(dx)$ n'est pas une mesure de probabilité. Dans le but de réécrire m_j en terme d'espérance, nous allons donc renormaliser notre mesure $f(x)\mu(dx)$:

$$m_j = \int_{\Omega} Z w_j(\bar{x}) \frac{f(\bar{x}) d\mu(\bar{x})}{Z}$$

avec

$$Z = \int_{\Omega} f(x) \mu(dx)$$

Nous allons chercher à simuler une chaîne de Markov ergodique (X_1, \dots, X_N) , selon la mesure invariante

$$p(dx) = \frac{f(x)\mu(dx)}{Z} \propto f(x)\mu(dx)$$

Remarques :

1. L'algorithme de Metropolis nécessite que la mesure invariante à simuler soit supposée charger tous les points de Ω . Or, à cause de la définition de G dans (3.1), la fonction f que nous avons définie peut être nulle. Par conséquent, pour utiliser l'algorithme de Metropolis, il faudra supposer que f ne s'annule pas ; c'est à dire restreindre l'ensemble Ω à l'ensemble des trajectoires de lumière possible.
2. Si $y_1 \in \mathcal{M}^{n_1}$ et $y_2 \in \mathcal{M}^{n_2}$, alors $\mu(\{y_1\}) = \mu(\{y_2\}) = 1$.

L'algorithme est le suivant. On se donne un noyau de transition $Q : \Omega \times \Omega \rightarrow [0, 1]$, irréductible, et tel que $Q(x, y) > 0 \Rightarrow Q(y, x) > 0$ pour tout $x, y \in \Omega$

Initialisation : On 'tire' X_0 selon une mesure de Dirac $\delta_x, x \in \Omega$

Puis pour $n = 1, \dots, N - 1$

$$\left\{ \begin{array}{l} \text{On tire } u \text{ suivant la loi uniforme } \mathcal{U}([0, 1]), y \text{ dans } Q(X_n, \cdot) \\ \text{On calcule } \alpha(x_n, y) = \min \left\{ 1, \frac{f(y)\mu(\{y\})Q(y, x_n)}{f(x_n)\mu(\{x_n\})Q(x_n, y)} \right\} \\ \text{Si } u < \alpha(x_n, y) \text{ On accepte } y : X_{n+1} = y. \text{ Sinon, on rejette } y : X_{n+1} = X_n \end{array} \right.$$

Alors si l'hypothèse d'ergodicité de la chaîne est vérifiée pour ce choix de Q (par exemple $Q(x, y) > 0$ pour tout $x, y \in \Omega$), un estimateur consistant et sans biais de m_j , lorsque N est suffisamment grand est donnée par :

$$\hat{m}_j = \frac{1}{N} \sum_{k=0}^{N-1} Z w^{(j)}(X_k). \quad (3.8)$$

De plus pour n suffisamment grand, on aura $\mathcal{L}(X_n)$ proche de p .

Nous pouvons donc construire un estimateur consistant \tilde{m}_j de $\frac{m_j}{Z}$ défini par :

$$\tilde{m}_j = \frac{1}{N} \sum_{k=0}^{N-1} w^{(j)}(X_k).$$

Dans la partie suivante, nous allons parler de la stratégie adoptée pour le choix du noyau Q . En effet, il est important de bien spécifier Q duquel dépend l'ergodicité de la chaîne, ainsi que le taux d'acceptation.

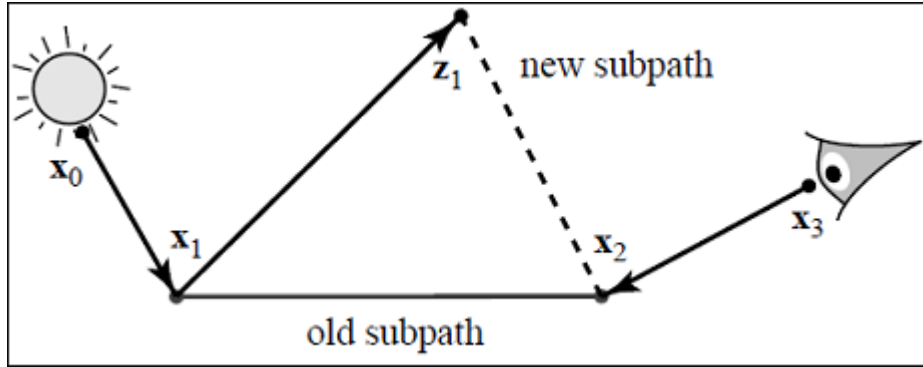


FIGURE 3.1 – [VEACH 97] : Un exemple de mutation bidirectionnelle ; le chemin $x_0x_1x_2x_3$ est remplacé par le chemin $x_0x_1z_1x_2x_3$.

3.4 Mutations bidirectionnelles

Nous allons adopter une stratégie bidirectionnelle dans la construction du noyau Q de l'algorithme de Metropolis, que nous allons détailler ci-après.

Notons $\bar{x}_c = x_0 \dots x_s x_{s+1} \dots x_{t-1} x_t \dots x_k$ le chemin courant et $\bar{x}_n = x_0 \dots x_s x_{l_1} \dots x_{l_p} x_t \dots x_k$ le nouveau chemin, dont la taille n'est plus forcément identique.

Ainsi, $Q(\bar{x}_c, \bar{x}_n)$ sera la probabilité que la nouvelle trajectoire soit égale à \bar{x}_n , sachant que l'ancienne était \bar{x}_c . En d'autres termes, ce sera la probabilité de supprimer la partie $x_s \dots x_t$, trajectoire de taille $l_d = t - s$ (la taille est définie par le nombre de ponts qui forment deux sommets) et de la remplacer par $x_s x_{l_1} \dots x_{l_p} x_t$, trajectoire de taille l_a .

Tirer \bar{x}_n selon la loi $Q(\bar{x}_c, \cdot)$ se fait donc en deux étapes :

1. – On tire une taille aléatoire $l_d = t - s$ grâce à une loi $p_{d,1}$
 - On tire un entier s aléatoirement grâce à une loi $p_{d,2}$
 - On obtient $t = l_d + s$ et on supprime le chemin menant de x_s à x_t de la trajectoire initiale.
2. – On tire une nouvelle taille $l_a = s' + t' + 1$ suivant une loi $p_{a,1}$
 - On tire un entier s' uniformément sur l'ensemble des valeurs que peut prendre s' en fonction de la taille l_a choisie
 - On obtient un $t' = l_a - s' - 1$. s' (resp. t') correspond au nombre de nouveaux sommets à ajouter après x_s (resp. avant x_t). C'est ce point qui justifie l'appellation de mutation bidirectionnelle.
 - On tire $(x_{l_1})_{l_1=1, \dots, s'}$, x_1 selon la loi $\hat{f}(x_{s-1}, x_s, \cdot)G(x_s, \cdot)$, x_2 selon la loi $\hat{f}(x_s, x_{l_1}, \cdot)G(x_{l_1}, \cdot)$, etc. Et de la même façon de l'autre côté on tire $(x_{l_2})_{l_2=1, \dots, t'}$.
3. On construit ainsi notre nouveau chemin $\bar{x}_n = x_0 \dots x_s x_{l_1} \dots x_{l_p} x_t \dots x_k$.

Nous noterons $p_d[s, t] = p_{d,1}p_{d,2}$ la probabilité de supprimer la sous trajectoire $x_s \dots x_t$ et $p_a[s', t'] = p_{a,1}p_{a,2}$ la probabilité de rajouter s' sommets après x_s et t' sommets avant x_t .

Par exemple : supposons que $\bar{x}_c = x_0x_1x_2x_3$ et que nous avons pioché $y = \bar{x}_n = x_0x_1x_2x_3$, (figure 3.4) en ayant obtenu de façon aléatoire $l_d = 1, s = 1, (t = 2), l_a = 2, s' = 1 (t' = 0)$. Ainsi :

$$Q(\bar{x}_c, \bar{x}_n) = p_d[1, 2] [p_a[1, 0]f_s(x_0, x_1, x_{l_1})G(x_1, x_{l_1}) + p_a[0, 1]f_s(x_3, x_2, x_{l_1})G(x_2, x_{l_1})].$$

De la même façon, on a :

$$Q(\bar{x}_n, \bar{x}_c) = p_d[1, 3]p_a[0, 0].$$

Ainsi, après simplification des termes $L_e(x_0, x_1)$, $G(x_0, x_1)$ et $G(x_2, x_3)$, on obtient :

$$\alpha(\bar{x}_c, \bar{x}_n) = \frac{f_s(x_0, x_1, x_{l_1})G(x_1, x_{l_1})f_s(x_1, x_{l_1}, x_2)G(x_{l_1}, x_2)f_s(x_{l_1}, x_2, x_3)}{f_s(x_0, x_1, x_2)G(x_1, x_2)f_s(x_1, x_2, x_3)} \frac{Q(\bar{x}_n, \bar{x}_c)}{Q(\bar{x}_c, \bar{x}_n)}.$$

Dans [VEACH 97], il est suggéré de prendre pour $p_{d,1}$:

$$\begin{aligned} \mathbb{P}(l_d = 2) &= 0.5, \\ \mathbb{P}(l_d = 1) &= 0.25, \\ \mathbb{P}(l_d = k) &= 2^{-k} \text{ pour } k \geq 3. \end{aligned}$$

De même, à l_d fixé, pour $p_{a,1}$:

$$\begin{aligned} \mathbb{P}(l_a = l_d) &= 0.5, \\ \mathbb{P}(l_a = l_d + 1) &= \mathbb{P}(l_a = l_d - 1) = 0.15, \\ \mathbb{P}(l_a = l_d + j) &= \mathbb{P}(l_a = l_d - j) = 0.2(2^{-j}) \text{ pour } j \geq 2. \end{aligned}$$

Remarque :

De part la construction de Q , à partir de n'importe quel état on peut arriver à n'importe quel autre état puisque on peut choisir (avec faible probabilité) de supprimer l'intégralité du chemin et de le remplacer par un autre [$q_{ij} > 0 \forall i, j \in \Omega$]. Par conséquent, la chaîne engendrée par l'algorithme de Metropolis est ergodique de matrice de transition irréductible.

Dans ce chapitre, nous avons repris les algorithmes pour la synthèse d'image décrits dans [VEACH 97] et [PAULY 00]. Le chapitre suivant sera une mise en oeuvre des algorithmes MLT initialement prévus pour la synthèse d'image, dans le cas de la propagation de la lumière dans un tissu du cerveau que l'on supposera homogène, soit, en l'absence de tumeur. Les tissus du cerveau étant caractérisés par des coefficients de diffusion, d'absorption et d'anisotropie, nous ne pourrions pas négliger l'influence du milieu sur la dynamique de la lumière, ainsi nous reprendrons la stratégie adopté dans 3.2, en l'absence de surface (absence de points frontière).

Chapitre 4

Propagation de la lumière dans le cerveau

Nous allons voir dans ce chapitre, comment mettre en place l'algorithme Métropolis Light Transport décrit dans le chapitre précédent, pour simuler la diffusion de la lumière un tissu du cerveau, considéré dans le cadre de ce mémoire, comme étant un **milieu homogène**, c'est à dire un milieu dans lequel les coefficients de diffusion, d'absorption et d'anisotropie ne varient pas et donc où il n'y a pas de tumeur.

On se donne un coefficient d'anisotropie $g=0.9$, un coefficient de diffusion $\sigma_s = 280cm^{-1}$ et un coefficient d'absorption $\sigma_a = 0.57cm^{-1}$, obtenus à partir d'expérience sur le rat, que l'on pourra trouver dans la littérature, voir [ANGEL-PETERSEN 07].

À terme, nous allons résoudre le problème inverse qui consistera à estimer ces coefficients à partir d'expériences, de façon à ce que la diffusion de la lumière dans les tissus que l'on obtiendra collera au mieux à la réalité.

4.1 Mise en place de l'algorithme

Afin de déterminer la quantité de lumière perçue par chaque points j de notre espace que constitue le tissu d'étude de notre cerveau, nous allons chercher à estimer les m_j décrits dans 3.3 grâce aux estimateurs 3.8. Comme choix du noyau de transition Q , nous allons adopter la stratégie bidirectionnelle décrite dans la partie 3.4. Il nous reste à décrire la fonction de phase f_p utilisée dans notre cas, ainsi que l'obtention de nouveaux sommets.

Sachant un sommet x_i , pour choisir un nouveau sommet x_{i+1} nous avons besoin d'une direction w et d'une distance d . La distance sera tiré dans une loi exponentielle de paramètre $\sigma_e = \sigma_a + \sigma_s$ [PAULY 00].

Une direction dans \mathcal{S}^2 peut être définie par deux coordonnées sphériques θ et ϕ . Tirer une direction ω dans \mathcal{S}^2 se fait donc en deux étapes. On tire

aléatoirement les angles sphériques θ et ϕ selon la distribution choisie, puis on effectue un changement de base pour obtenir le vecteur directionnel $w = [w_x \ w_y \ w_z]$ en coordonnées cartésiennes :

$$\begin{cases} w_x = \sin(\theta) \cos(\phi) \\ w_y = \sin(\theta) \sin(\phi) \\ w_z = \cos(\theta) \end{cases} .$$

Nous obtenons alors les coordonnées du sommet suivant : $x_{i+1} = x_i + d.w$.

Nous placerons le premier sommet, point de diffusion de la lumière sur la fibre optique, à l'origine du repère. La fibre étant caractérisée par un angle de demi ouverture α (voir figure 4.1.1), l'angle sphérique θ appelé angle de déflexion, sera choisi uniformément entre $\pi - \alpha$ et π . L'angle azimutal ϕ sera alors choisi uniformément entre 0 et 2π .

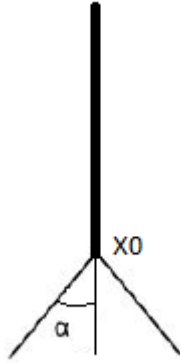


FIGURE 4.1 – Fibre

Pour obtenir les sommets suivants, l'angle azimutal ϕ sera à nouveau choisi uniformément entre 0 et 2π . Cependant nous tirerons le cosinus de l'angle θ aléatoirement dans la fonction de phase d'Henry-Greenstein, définie dans [WANG 95] par :

$$f_p(y) = \frac{1 - g^2}{2(1 + g^2 - 2gy)^{\frac{3}{2}}} \mathbb{1}_{[-1,1]}(y),$$

où g est le coefficient d'anisotropie de notre milieu. (On pourra aisément vérifier que f_p est bien une densité.)

Pour cela, nous allons utiliser la méthode de la fonction de répartition inverse. En effet, si F est la fonction de répartition de $Y = \cos(\theta)$, alors la loi de Y est la même que celle de $F^{\leftarrow}(\xi)$ où F^{\leftarrow} désigne l'inverse généralisée de F , ξ une variable uniforme sur $[0, 1]$.

Pour $g \neq 0$, on a :

$$F(y) = \int_{-1}^y f(x)dx = \frac{1-g^2}{2g(1+g^2-2gy)^{\frac{1}{2}}} - \frac{1-g^2}{2g(1+g)}.$$

D'où pour $g \neq 0$:

$$F^{-1}(\xi) = \frac{1}{2g} \left(1+g^2 - \left[\frac{1-g^2}{1-g+2g\xi} \right]^2 \right).$$

Ainsi :

$$Y = \cos(\theta) \stackrel{\mathcal{L}}{=} \begin{cases} \frac{1}{2g} \left(1+g^2 - \left[\frac{1-g^2}{1-g+2g\xi} \right]^2 \right) & \text{si } g \neq 0, \\ 2\xi - 1 & \text{sinon.} \end{cases}$$

les angles ϕ et θ étant sélectionnés, il faut alors calculer les coordonnées sphériques de la nouvelle direction. D'après la formule de changement de base des coordonnées cartésiennes vers les coordonnées sphériques, la nouvelle direction $w_2 = [w_{2,x} \ w_{2,y} \ w_{2,z}]$ est définie par les angles $\theta + \arccos(w_z)$ et $\phi + \arccos\left(\frac{w_x}{\sqrt{w_x^2 + w_y^2}}\right)$ (pour plus de détail, voir [WANG 95]). Ainsi :

$$\begin{cases} w_{2,x} = \sin(\theta + \arccos(w_z)) \cos\left(\phi + \arccos\left(\frac{w_x}{\sqrt{w_x^2 + w_y^2}}\right)\right) \\ w_{2,y} = \sin(\theta + \arccos(w_z)) \sin\left(\phi + \arccos\left(\frac{w_x}{\sqrt{w_x^2 + w_y^2}}\right)\right) \\ w_{2,z} = \cos(\theta + \arccos(w_z)) \end{cases} .$$

Nous allons décrire dans la partie suivante le code matlab utilisé dans la programmation de cet algorithme. On pourra retrouver le code matlab en Annexe.

4.2 Algorithme

On choisie une trajectoire de lumière $X_0 = x_0 x_1 \dots x_k$ initiale.

Puis pour $n = 1, \dots, N$:

$\left\{ \begin{array}{l} \text{On choisi un chemin } Y \text{ dans } Q(\cdot | X_n) : \\ \left\{ \begin{array}{l} \text{On tire } l_d \text{ le nombre d'arrêtes à supprimer selon une loi choisie.} \\ \text{On tire un entier } s \text{ uniformément entre } 1 \text{ et } n - l_d. \\ \text{On supprime la sous partie } x_{s+1} \dots x_{t-1}, t = l_d + s \text{ du chemin } X_n. \\ \text{On tire } l_a \text{ le nombre d'arrêtes à ajouter selon une loi choisie.} \\ \text{On tire un entier } s_2 \text{ uniformément entre } 0 \text{ et } l_a - 1 \text{ et on en déduit } t_2 = l_a - s_2 - 1. \\ \text{On tire } l_a - 1 \text{ nouveaux sommets grâce à la fonction } \textit{generephoton} \text{ ou } \textit{generephotoninit} \\ \text{et l'on place } s_2 \text{ sommets après } x_s \text{ et } t_2 \text{ avant } x_t. \end{array} \right. \\ \text{On calcule la probabilité } Q(Y | X_n) \text{ d'obtenir le chemin } Y. \\ \text{On calcule le rapport : } \alpha = \frac{Q(X_n | Y) f(Y)}{Q(Y | X_n) f(X)} \\ \text{On tire } U \sim \mathcal{U}([0, 1]). \text{ Si } U < \alpha \text{ alors } X_{n+1} = Y, \text{ sinon } X_{n+1} = X_n \\ \text{FIN} \end{array} \right.$

Puis pour chaque point j de l'espace, on estime les m_j en calculant le nombre moyen de chemins de notre chaîne X formée, finissant par j .

Sur la Figure 4.2, on peut constater que les points éclairés sont conformes à ce quoi on pourrait s'attendre. On voit très nettement le cône formé par l'angle d'ouverture de la fibre. De plus, les points les plus illuminés (en noir, puis en jaune, puis en bleu) se trouvent à proximité du point de diffusion de la lumière (fixé à l'origine, en rouge).

Pour cette simulation de un million de pas, il a fallu environ une demi heure de temps de calcul. 100000 pas prennent une à deux minutes. Nous avons obtenu pour cette réalisation un taux d'acceptation d'environ 40%.

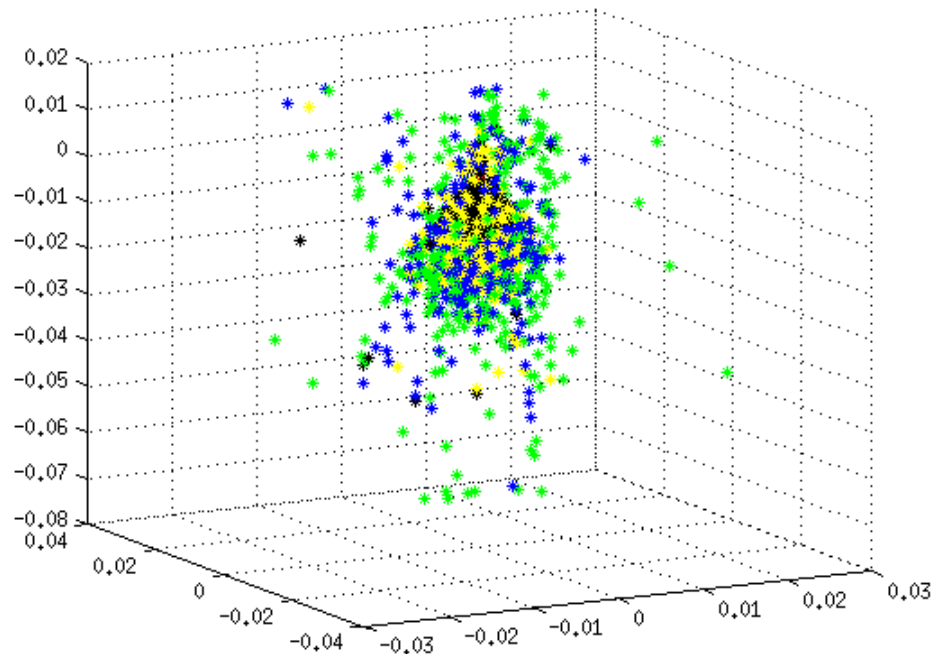


FIGURE 4.2 – Simulation de la propagation de la lumière dans un tissu supposée homogène, de paramètres $\sigma_a = 0.57cm^{-1}$, $\sigma_s = 280cm^{-1}$, $g = 0.9$ et pour un angle de demi ouverture de la fibre $\alpha = \pi/6$, avec $K=1000000$ pas. Sur ce graphique, on peut voir en rouge l'origine du repère, point de diffusion de la lumière. Les couleurs sont séparées selon les quartiles : les points noirs représentant les points les plus illuminés, puis les jaunes, les bleus et enfin les verts.

4.3 Conclusion

Dans ce premier travail, nous avons adapté les algorithmes décrits dans [VEACH 97] et dans [PAULY 00] pour simuler la lumière dans un tissu ho-

mogène. On a pu constater que les représentations graphiques que nous donne ces algorithmes sont proches de ce dont on pourrait s'attendre. Le temps de calcul est relativement faible (une minute pour 100000 simulations), temps qui pourra être optimisé en parallélisant le problème. Il faudra par la suite intégrer une tumeur, voir d'autres tissus du cerveaux, qui constitueront les frontières et pour lesquelles on utilisera la fonction de distribution bidirectionnelle f_s qui sera une distribution de Bernoulli prenant en paramètre les indices de réfractions (noté n_i et n_t sur la figure 4.3) propres à chacun des milieux. En effet, le rayon lumineux atteignant une frontière aura deux choix : il pourra être réfléchi selon le même angle θ_i angle entre le vecteur directionnel et la normale, ou réfracté dans le nouveau milieu selon un autre angle θ_t défini par les caractéristiques du milieu.

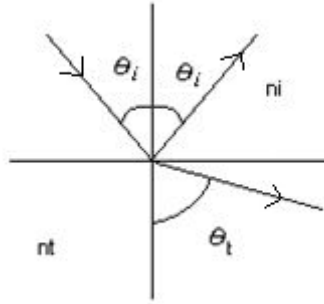


FIGURE 4.3 – Reflexion Refraction

Pour construire un nouveau sommet y_i à partir d'un chemin (y_0, \dots, y_{i-1}) , la stratégie sera alors la suivante :

1. On tire une direction $\omega \in \mathcal{S}^2$ dans la fonction de phase f_p si y_{i-1} est dans le volume, et dans la fonction de distribution bidirectionnelle f_s sinon.
2. On tire une direction d dans une loi exponentielle de paramètre σ_e
3. On calcule $z = y_{i-1} + \omega X$. Si le point z a dépassé une surface, alors on choisira $y_i = \pi_\omega(y_{i-1})$. Sinon on prendra $y_i = z$

Des expériences seront faites sur des rats, afin de comparer les résultats obtenus avec des résultats expérimentales. Ces expériences serviront également à la résolution du problème inverse, à savoir définir les paramètres de diffusion, d'absorption et d'anisotropie des différents milieux, de telle sorte que les résultats expérimentales collent au mieux avec les résultats obtenus par méthodes de Monte Carlo par Chaîne de Markov. Une étude sur le type de fibre optique à utiliser, ainsi que sur son placement optimale (de telle sorte à ce que la tumeur soit irradiée de la meilleure façon) se révélera également nécessaire.

Conclusion

Pour déterminer la dose de lumière émise par une fibre optique, reçue par une tumeur ainsi que sa répartition spatiale en fonction des caractéristiques de la source lumineuse, il est possible d'utiliser les lois de la physique en utilisant l'équation de transfert radiatif (ERT). Pour résoudre cette équation dans le cas de tissus biologiques, il est possible d'utiliser une représentation probabiliste de cette équation et d'utiliser une méthode de Monte Carlo de simulation de photons.

Dans ce mémoire, nous avons abordé une méthode alternative à la méthode de Monte Carlo par simulation de photons, pour simuler la diffusion de la lumière dans le cerveau. Elle réside en une adaptation des algorithmes pour la synthèse d'image communément appelés 'Metropolis Light Transport' dans la littérature. Les rayons simulés par cette méthodes perdent leur sens physique, mais on peut néanmoins retrouver une interprétation en terme de rayon lumineux. Cette méthode présente l'avantage sur la méthode de Monte Carlo, d'une très intéressante économie de temps de calcul et offre des perspectives de parallélisation.

Le but de ce premier chapitre, était la démonstration des théorèmes ergodiques, sur lesquels reposent les algorithmes de Monte Carlo par Chaînes de Markov (MCMC). Nous nous sommes basés sur le cas des chaînes de Markov homogènes à temps discret sur un espace d'états dénombrables. Ces théorèmes ergodiques sont également possibles en considérant des chaînes sur un espace d'états indénombrables, en considérant non plus des matrices de transition, mais des noyaux de transition qui seront des fonctions positives de $E \times E$ dans $[0, 1]$, telles que $\int_E Q(x, dy) = 1$ quelque soit $x \in E$. En effet, dans notre travail, nous avons estimé les valeurs des pixels j sur une grille de points que l'on considère dénombrable. L'ensemble des coordonnées de \mathbb{R}^3 étant infini, on pourrait essayer de se placer dans un espace d'états infini, plutôt que de travailler sur une grille de points.

Dans le second chapitre, nous avons étudié les algorithmes dit MCMC, plus particulièrement les algorithmes de Métropolis. Cependant dans notre travail, notre espace d'états Ω est en réalité l'union de plusieurs sous-espaces disjoints. Il existe des algorithmes plus adaptés au cas où l'espace d'états est composé de sous espaces, appelés *Reversible Jumping Rule*. On pourra les retrouver par exemple dans Jun S. Liu [LIU 08]. Leur analyse pourrait se révéler intéressante.

Une analyse de la vitesse de convergence des algorithmes est plus que nécessaire,

afin de pouvoir déterminer le nombre de simulation envisagées pour atteindre la loi cible. Dans un prochain travail, nous essayerons de voir s'il est possible de mettre en place la stratégie spectrale décrite dans le chapitre 2.

Par application du chapitre 3, nous avons pu simuler la lumière à travers un tissu homogène selon les caractéristiques de milieu d'un cerveau de rat. Nous avons pu constater que le flux de la lumière à travers le tissu est conforme à ce qu'on pourrait s'attendre. Cependant, dans ce mémoire, aucune preuve n'est donnée concernant la démonstration de l'hypothèse : $\|T\| < 1$ sur laquelle repose l'algorithme. On pourra s'appuyer sur [PAULY 00] pour obtenir notre preuve. Parallelement, il faudra intégrer à présent une tumeur au problème ainsi que d'autres couches du cerveau.

Deux autres perspectives à la fin de ces travaux seront d'une part d'optimiser le placement de la fibre (qui peut être à l'intérieur même de la tumeur), de telle sorte que la tumeur soit irradiée au mieux. De même, une étude comparative sur le type de fibre optique en fonction de la lumière qu'elle diffuse devra être envisagée. L'autre axe de recherche sera de résoudre le problème inverse : c'est à dire déterminer de nous même les coefficients d'anisotropies, d'absorptions et de diffusions des différents milieux du cerveau de telle sorte que la diffusion de la lumière obtenue par expérience colle au mieux à la diffusion obtenue par simulation de Monte Carlo par Chaînes de Markov.

Bibliographie

- [ANGEL-PETERSEN 07] E. Angell-Petersen, H. Hirschberg, and S. J. Madsen : *Determination of fluence rate and temperature distributions in the rat brain ; implications for photodynamic therapy*. Journal of biomedical optics, 12 (1), 2007.
- [BENAIM 05] M. Benaïm, N. El Karoui : *Promenade aléatoire : Chaînes de Markov et simulations ; martingales et stratégies*. Editions l'Ecole Polytechnique, 2005.
- [GRIMMETT 01] Grimmett, Geoffrey R. and Stirzaker, David R : *Probability and random processes*. Third Edition. Oxford University Press, New York, 2001.
- [LIU 08] Jun S. Liu : *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer, New York, 2008.
- [LACROI 05] J. Lacroix , P. Priouret *Cours de probabilités approfondies* , Jussieu : Université Pierre et Marie Curie, 2005. pp. 73-98 Disponible sur [http ://www.proba.jussieu.fr/cours/processus.pdf](http://www.proba.jussieu.fr/cours/processus.pdf) (consulté le 21 juin 2011)
- [PAULY 00] M. Pauly, T. Kollig, A. Keller : Metropolis light transport for participating media. *Rendering Techniques 2000 (Proc. of Eurographics Rendering Workshop)*, pp. 11-22, 2000.
- [PEKSUN 73] P.H. Peskun *Optimum Monte-Carlo Sampling Using Markov Chains*. Biometrika, Vol. 60, No. 3. pp. 607-612, 1973
- [ROBERT 96] C. Robert : *Méthodes de Monte Carlo par chaînes de Markov*. Méthodes de Monte Carlo par chaînes de Markov. (French. French summary) [Markov chain Monte Carlo methods] Statistique Mathématique et Probabilité. [Mathematical Statistics and Probability] Éditions Économica, Paris, 1996.
- [VEACH 97] E. Veach, L. Guibas : Metropolis light transport. *Proc. of SIGGRAPH 1997*, p. 419-428, 1997.
- [WANG 95] L. Wang , Steven L. Jacques and Liqiong Zheng : MCML-Monte Carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedecine*, 1995 n0. 47, pp. 131-146

Annexe A

Codes R du chapitre 2

A.1 loi de Poisson

```
pois_sample<-function(n,x_0,l)
{
  C<-rep(0,n)
  C[1]<-x_0
  for(i in 1:n)
  {
    u<-runif(1,0,1)
    if(C[i]==0 & u<1/2){C[i+1]<-1}
    if(C[i]==0 & u>1/2){C[i+1]<-0}

    else
    {
      u<-runif(1,0,1)
      y<-runif(1,0,1)
      if(y<1/2 & u<C[i]/l){C[i+1]<-C[i]-1}
      if(y<1/2 & u>C[i]/l){C[i+1]<-C[i]}
      if(y>1/2 & u<1/C[i]){C[i+1]<-C[i]+1}
      if(y>1/2 & u>1/C[i]){C[i+1]<-C[i]}
    }
  }
  par(mfrow=c(1,2))
  hist(C,main="Simulation par algorithme de Metropolis",xlab="",ylim=c(0,2000),xlim=c(0,15))
  hist(rpois(n,l),main="Simulation exacte",xlab="",ylim=c(0,2000),xlim=c(0,15))
  C
}
```

A.2 Modèle des sphères dures

```
sphere_dure<-function(S,n)
{
  X<-rep(0,S+2) #On augmente la taille de 2 pour éviter les problème de frontière
  M<-matrix(rep(0,n*(S+2)),n,S+2)
```



```
C<-rep(0,S)
for(i in 1:n)
{
s<-sample(2:(S+1))[1]#On tire un entier au hasard entre 2 et S+1
while(X[s+1]==1 | X[s-1]==1)
{s<-sample(2:(S+1))[1]}

u<-runif(1,0,1)
if(u<1/2){X[s]<-1}
else {X[s]<-0}
M[i,]<-X
}
M<-M[,-(S+2)] #On prend garde d'oter les 2 colonnes ajoutés
M<-M[,-1]

for(i in 1:n){C[i]<-sum(M[i,])}

plot(C,xlab="Etape",ylab="Nombre de particules presentes" )
abline(h=mean(C),col="red")
mean(C)
}

sphere_dure(200,10000)
>54.5983
```

Annexe B

codes matlab du chapitre 4

B.1 Génération du premier sommet :

```
function Z=genere_photon_init(sig_s,sig_a,g,X,alpha,l)
X=X(:);
cos_theta=-1+rand(1)*(cos(pi-alpha)+1);
theta=acos(cos_theta);
phi=2*pi*rand(1);
d=exprnd(1/(sig_s+sig_a));
x2=sin(theta)*cos(phi);
y2=sin(theta)*sin(phi);
z2=cos(theta);
w=[x2,y2,z2]';
Z=X+d*w;
```

B.2 Génération des sommets suivants :

```
function Z = genere_photon(sig_s, sig_a, g, X, Y,l)
X=X(:);
Y=Y(:);
w=(Y-X)/norm(X-Y);
theta=sample_phase(g);
phi=2*pi*rand(1);
d=exprnd(1/(sig_s+sig_a));
if(arrondi(w(3),4)==1 || arrondi(w(3),4)==-1)
    xw=sin(theta)*cos(phi);
    yw=sin(theta)*sin(phi);
    zw=sign(w(3))*w(3)*cos(theta);
else
    sintheta=sqrt(1-w(3)^2)*cos(theta)+w(3)*sin(theta);
    costheta=w(3)*cos(theta)-sqrt(1-w(3)^2)*sin(theta);
    cosphi=(w(1)*cos(phi)-w(2)*sin(phi))/sqrt(1-w(3)^2);
    sinphi=(w(1)*sin(phi)+w(2)*cos(phi))/sqrt(1-w(3)^2);
    xw=sintheta*cosphi;
    yw=sintheta*sinphi;
```

```

        zw=costheta;
    end
    w=[xw,yw,zw]';

    Z=Y+d*w;

```

Où générer un angle θ dans la fonction de phase se fait grâce à la fonction :

```

function theta=sample_phase(g)
cos=(1+g^2-((1-g^2)/(1-g+2*g*rand(1)))^2)/(2*g);
theta=acos(cos);

```

B.3 Densité de probabilité d'obtenir un nouveau sommet

```

function p=dens_photon(sig_s,sig_a,g,X,Y,Z)
u=Y-X;
v=Z-Y;
cos_theta=(sum(u.*v)/(norm(Y-X)*norm(Z-Y)));
theta=acos(cos_theta);
d=norm(Z-Y);
p=f_p(g,cos(pi-theta))*(sig_s+sig_a)*exp(-d*(sig_s+sig_a))/(2*pi);

```

Où f_p est la fonction de phase définie par :

```

function x=f_p(g, costheta)
x=(1-g^2)/(2*(1+g^2-2*g*costheta)^(3/2));

```

B.4 Choix du nombre de pont l_d à supprimer

```

function ld=sample_ld()
e=rand(1);
ld=2;
x=1/4;
if(e>1/2)
    ld=2;
else if(e<1/2 && e>1/4)
    ld=1;
    else while(e>x || e<x/2)
        ld=ld+1;
        x=x/2;
    end
end
end

```

B.5 Probabilité d'obtenir un l_d donné

```

function pd1=pd(ld,n)
if(ld==1), pd1=1/4;

```

B.6. CHOIX DU NOMBRE DE PONT L_A À AJOUTER, EN FONCTION DE L_D 67

```
else if(ld==2), pd1=1/2;
else pd1=1/2^(ld);
    end
end
end
```

B.6 Choix du nombre de pont l_a à ajouter, en fonction de l_d

```
function la=sample_la(ld)
e=rand(1);
b=rand();
x=0.1;
if(e>0.5), la=ld; end
if(e<0.5 && e>0.45), la=ld+1; end
if(e<0.45 && e>0.10), la=ld-1; end
if(b>0.8 && e<0.1)
    la=ld+2;
    while(e>x || e<x/2)
        la=la+1;
        x=x/2;
    end
end
if(b<0.8 && e<0.1)
    la=ld-2;
    while(e>x || e<x/2)
        la=la-1;
        x=x/2;
    end
end
if(la<1)
    la=1;
end
```

B.7 Probabilité d'obtenir un l_a donné conditionnellement à l_d

```
function pa2=pa(ld,la,t,n)
if(la==ld), pa1=0.5;
end
if(la==ld+1), pa1=0.05; end
if(la==ld-1), pa1=0.35; end
if(la>ld+1) pa1=0.2*0.1*2^(-abs(la-ld)); end
if(la<ld-1) pa1=0.8*0.1*2^(-abs(la-ld)); end
```

B.8 Echantillonner un nouveau chemin dans la loi Q , sachant un chemin X_0

```

function [M,s,t,ld,la]=sample_Q(sig_s,sig_a,g,X0,alpha,l)
n=size(X0,2);
ld=sample_ld();
if(ld>n)
    ld=n;
end
s=round((n-ld)*rand()+1);
t=ld+s;
la=sample_la(ld);

if(t<=n)
s2=round((la-1)*rand()) ;
t2=la-s2-1;
else
    s2=la-1;
    t2=0;
end
M=[];
for i=1:s
    M(:,i)=X0(:,i);
end
if(t<=n)
    for i=t:n
        M(:,i-ld+la)=X0(:,i);
    end
end

if(s2>0)
    if(s==1), M(:,s+1)=genere_photon_init(sig_s,sig_a,g,M(:,s),alpha,l);
    else M(:,s+1)=genere_photon(sig_s,sig_a,g,M(:,s-1),M(:,s),l); end
    if(s2>1)
        for i=2:s2 , M(:,s+i)=genere_photon(sig_s,sig_a,g,M(:,s+i-2),M(:,s+i-1),l); end
    end
end

if(t2>0)
    if(t==n), M(:,t-1-ld+la)=genere_photon_init(sig_s,sig_a,g,M(:,t-ld+la),alpha,l);
    else M(:,t-1-ld+la)=genere_photon(sig_s,sig_a,g,M(:,t+1-ld+la),M(:,t-ld+la),l);
    end
    if(t2>1)
        for j=2:t2
            M(:,t-j-ld+la)=genere_photon(sig_s,sig_a,g,M(:,t-j+2-ld+la),M(:,t-j+1-ld+la),l)
        end
    end
end

if(size(M,2)==1), [M,s,t,ld,la]=sample_Q(sig_s,sig_a,g,X0,alpha,l); end

```

B.9 calcul de la probabilité d'obtenir ce nouveau chemin

```
function p=Q(sig_s,sig_a,g,X,Y,alpha,ld,la,s,t)
n=size(X,2)
if(t>n) pa2=1;
else pa2=1/la;
pd2=1/(n-ld+1);

h=0;
if(la==1), h=pa(ld,la,t,n);
else
    for i=0:(la-1)
        h=h+pa2*pa(ld,la,t,n)*pg(sig_s,sig_a,g,s,t,i,la-i-1,ld,la,X,Y,alpha);
    end
end
p=h*pd2*pd(ld,n);
```

où pg est le produit des densités :

```
function h=pg(sig_s,sig_a,g,s,t,s2,t2,ld,la,X,Y,alpha)
h=1;
if(s2>0)
    if(s==1)
        h=h*(1/(2*pi*(cos(pi-alpha)+1)));
        if(s2>1)
            for i=2:s2
                h=h*dens_photon(sig_s,sig_a,g,Y(:,s+i-2),Y(:,s+i-1),Y(:,s+i));
            end
        end
    end
    if(s>1)
        for i=1:s2 h=h*dens_photon(sig_s,sig_a,g,Y(:,s+i-2),Y(:,s+i-1),Y(:,s+i)); end
    end
end

if(t2>0)
    if(t==size(X,2))
        h=h*(1/(2*pi*(cos(pi-alpha)+1)));
        if(t2>1)
            for j=2:t2
                h=h*dens_photon(sig_s,sig_a,g,Y(:,t-(ld-la)-j+2),Y(:,t-(ld-la)-j+1),Y(:,t-(ld-la)-j));
            end
        end
    end
    if(t<size(X,2))
        for j=1:t2
            h=h*dens_photon(sig_s,sig_a,g,Y(:,t-(ld-la)-j+2),Y(:,t-(ld-la)-j+1),Y(:,t-(ld-la)-j));
        end
    end
end
```

```
end
```

B.10 calcul de f

```
function tau=G(sig_s,sig_a,X,Y)
d=norm(Y-X);
tau=exp(-d*(sig_s+sig_a));
```

```
function x=hat_f(sig_s,g,X,Y,Z)
u=Y-X;
v=Z-Y;
cos_theta=(sum(u.*v)/(norm(Y-X)*norm(Z-Y)));
theta=acos(cos_theta);
x=sig_s*f_p(g,cos(pi-theta));
```

```
function h=f(sig_s,sig_a,g,X)
h=G(sig_s,sig_a,X(:,1),X(:,2));
if(size(X,2)>2)
    for i=2:(size(X,2)-1)
        h=h*hat_f(sig_s,g,X(:,i-1),X(:,i),X(:,i+1))*G(sig_s,sig_a,X(:,i),X(:,i+1));
    end
end
```

B.11 Algorithme de Metropolis

```
function [Z,taux_acceptation]=metropo(sig_s,sig_a,g,X0,N,K,alpha,l)
X=initialisation(sig_s,sig_a,g,X0,N,alpha,l);
X=arrondi(X,l);

    grid on
    hold on

test=[];
Z2=[];
Z=ones(4,K);
for m=1:3, Z(m,1)=X(m,size(X,2)-1); end
Z(4,1)=1;
for i=2:K

    [Y,s,t,ld,la]=sample_Q(sig_s,sig_a,g,X,alpha,l);
    Q_YX=Q(sig_s,sig_a,g,Y,X,alpha,la,ld,s,t+la-ld)
    Q_XY=Q(sig_s,sig_a,g,X,Y,alpha,ld,la,s,t)
    alpha2=min(1,(Q_YX*f(sig_s,sig_a,g,Y))/(Q_XY*f(sig_s,sig_a,g,X)));
    U=rand(1);
    if(U<alpha2)
        X=Y ;
        X=arrondi(X,l);
```

```

        test(i-1)=1;
    else
        test(i-1)=0;
    end

    Z2=Z;
    for j=1:(i-1)
        if(Z(1,j)==X(1,size(X,2)-1) && Z(2,j)==X(2,size(X,2)-1) && Z(3,j)==X(3,size(X,2)-1))
            Z(4,j)=Z(4,j)+1;
        end
    end
    if(Z2==Z)
    for m=1:3, Z(m,i)=X(m,size(X,2)-1); end
        Z(4,i)=1;
    end

end
taux_acceptation=sum(test)/(K-1)

J=[];
n=1;
x=1;

while n<=size(Z,2)
    if(Z(:,n)~=ones(4,1))
        J(:,x)=Z(:,n);
        x=x+1;
    end
    n=n+1;
end
J(4,:)=J(4,+)/K;
Z=J;

for i=1:size(Z,2)
    if(Z(4,i)<quantile(Z(4,:),0.25))
        plot3(Z(1,i),Z(2,i),Z(3,i),'*g')
    end
    if (Z(4,i)>quantile(Z(4,:),0.25) && Z(4,i)<quantile(Z(4,:),0.5))
        plot3(Z(1,i),Z(2,i),Z(3,i),'*b')
    end
    if (Z(4,i)>quantile(Z(4,:),0.5) && Z(4,i)<quantile(Z(4,:),0.75))
        plot3(Z(1,i),Z(2,i),Z(3,i),'*y')
    end

    if (Z(4,i)>quantile(Z(4,:),0.75))
        plot3(Z(1,i),Z(2,i),Z(3,i),'*black')
    end
end

plot3(0,0,0,'*r')

```