# Mémoire de Magistère de Mathématiques

Marie-Liesse Cauwet

Université Paris Sud XI
Orsay
2009-2013

# Table des matières

**1 Introduction** **2**

**2 Première année** **2**
2.1 Cours . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
2.2 Projet . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
2.3 Stage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

**3 Deuxième année** **5**
3.1 Cours . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
3.2 TER . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

**4 Année de préparation à l'agrégation** **6**

**5 Troisième année** **7**
5.1 Cours . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
5.2 Stage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

**6 Présentation de mon sujet de thèse** **8**
6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 8
6.2 Modélisation du problème : Prise de Décision Séquentielle . . 9
6.2.1 Pas de temps . . . . . . . . . . . . . . . . . . . . . . . 10
6.2.2 Variables d'état . . . . . . . . . . . . . . . . . . . . . . 10
6.2.3 Variables d'action . . . . . . . . . . . . . . . . . . . . . 10
6.2.4 Processus aléatoire . . . . . . . . . . . . . . . . . . . . 11
6.2.5 Fonction de transition . . . . . . . . . . . . . . . . . . 12
6.2.6 Fonction de récompense . . . . . . . . . . . . . . . . . 12
6.3 Méthodes de résolution du PPDS . . . . . . . . . . . . . . . 12
6.3.1 Politiques . . . . . . . . . . . . . . . . . . . . . . . . . 12
6.3.2 Fonction Valeur . . . . . . . . . . . . . . . . . . . . . . 13
6.3.3 Recherche Directe de Politique . . . . . . . . . . . . . 13
6.3.4 Équations de Bellman . . . . . . . . . . . . . . . . . . 14
6.4 Recherche Directe de Valeur . . . . . . . . . . . . . . . . . . 15
6.5 Développements envisagés : quelques pistes . . . . . . . . . . 15
6.5.1 Algorithmes d'optimisation . . . . . . . . . . . . . . . 15
6.5.2 Bagage probabiliste . . . . . . . . . . . . . . . . . . . 16

**7 Conclusion** **17**

**Références** **17**

# 1 Introduction

À l'issue de deux années en classe préparatoire, filiaire MP, j'ai passé les traditionnels concours d'entrée aux grandes écoles, mais je ne souhaitais pas poursuivre une formation d'ingénieur. C'est pourquoi j'ai choisi de continuer des études universitaires au sein du Magistère de Mathématiques d'Orsay. J'ai découvert l'existence du Magistère de Mathématiques d'Orsay grâce à une amie qui hésitait entre le Magistère de Mathématiques et celui de Physique, elle a opté pour la physique, et pour ma part, après avoir comparé celui de Rennes et celui d'Orsay, j'ai choisi ce dernier. En effet, la formation rigoureuse proposée, le large panel de cours dispensés, le bon niveau international du département de Mathématiques (bien situé au classement de Shanghai) ainsi que le bouche-à-oreille m'ont convaincue.

Dans le présent mémoire, je vais donc tenter de synthétiser ces trois années de Magistère, les cours que j'ai suivis, les projets que j'ai effectués, en mettant l'accent sur ce que le Magistère m'a apporté sur le plan des savoirs, mais aussi des différentes compétences que j'ai pu acquérir tout au long de ces études : sérieux, autonomie, rigueur. J'essayerai de montrer comment mon parcours au sein du Magistère a été déterminé par un mélange de facteurs tels que mes prédispositions pour telle ou telle branche des mathématiques ou mes goûts personnels.

Je décrirai les différents projets de recherche effectués pendant ces trois années de Magistère, sans toutefois restituer les rapports rédigés à chaque fois dans un souci de synthèse et afin d'apporter une vue globale de mon parcours plutôt qu'une succession de rapports. Je tiens néanmoins ces différents rapports à disposition (mail : marieliesse.cauwet@gmail.com).

Enfin, je mettrai l'accent sur la présentation de mon sujet de thèse, en détaillant le contexte général dans lequel elle s'inscrit, puis en présentant la modélisation adoptée et en signalant les améliorations et les pistes envisagées.

# 2 Première année

## 2.1 Cours

Les cours que j'ai suivis au sein de la L3 MFA durant cette première année de Magistère étaient les suivants :

- Topologie et Calcul différentiel ;
- Algèbre ;
- Intégration de Lebesgue ;

– Anglais ;
– Théorie des Équations différentielles ordinaires ;
– Analyse de Fourier ;
– Probabilités ;
– Fonctions holomorphes ;
– Algèbre effective et calcul formel.

Lors de cette première année de Magistère de mathématiques, trois cours supplémentaires nous étaient proposés en plus du parcours classique MFA. Au premier semestre, j'ai donc suivi le cours de Programmation, Algorithmique, Théorie de la complexité ainsi que le cours de Graphes et algèbre linéaire. Au second semestre, j'ai suivi le cours Complément de topologie et théorie de la mesure.

## 2.2 Projet

Mon projet de recherche a été réalisé sous l'encadrement de M. Guy Henniart. Il s'agissait d'étudier l'article [1] intitulé "PRIMES is in P" de Manindra Agrawal, Neeraj Kawal et Nitin Saxena, qui présente un algorithme déterministe permettant de déterminer en un temps polynomial si un nombre entier est premier.

**Définition 1** (Nombre premier)**.** *Un entier $n$ est dit premier si il admet exactement deux diviseurs distincts entiers et positifs, qui sont alors 1 et lui-même.*

On connaît depuis l'Antiquité, avec le crible d'Ératosthène, un algorithme capable de déterminer si un nombre est premier (cf Algorithme 1).

**Algorithm 1** Crible d'Ératosthène

**Entrée :** entier $n$
$m \leftarrow 2$
réponse $\leftarrow$ EstPremier
**Tantque** $m \leq \sqrt{n}$ et réponse = EstPremier
**Si** $m$ **divise** $n$
réponse $\leftarrow$ NonPremier
**FinSi**
**FinTantque**
**Retourner** réponse

Cependant, cette méthode, requiert $O(\sqrt{n})$ étapes pour déterminer si $n$ est premier, ce qui en fait un algorithme inefficace pour les grands nombres, alors même que connaître les grands nombres premiers se révèle extrêmement utile, en cryptographie par exemple. On aimerait donc disposer d'un algorithme de détermination de la primalité d'un nombre $n$ en un temps polynamial, i.e. en un nombre d'iterations en $O\left(log(n)\right)$.

L'article présente dans un premier temps l'algorithme [1] et fait la preuve de son exactitude, puis montre que cet algorithme est bien un algorithme en temps ploynomial et enfin propose des améliorations possibles. Pour ce projet, j'étais chargée de présenter la preuve que l'algorithme détecte bien les nombres premiers, alors qu'un autre étudiant analysait le temps de calcul. Les outils nécessaires à cette preuve sont issus de l'arithmétique modulaire et de la théorie des groupes (petit théorème de Fermat, polynômes cyclotomiques principlalement).

En rédigeant ce mémoire, je me suis aperçue qu'on retrouve dans tous les sujets de recherche que j'ai effectués, y compris celui-ci, des traits communs tels que l'application à des problèmes concrets, ou encore l'utilisation de l'outil informatique pour l'illustration ou pour le traitement de données. De ce point de vue là, ma thèse s'inscrit dans la continuité de mon parcours, bien que je sois désormais inscrite à l'école doctorale d'informatique.

### 2.3 Stage

Au cours de notre première année de Magistère, nous devions effectuer un apprentissage hors murs, sous la forme d'un stage en entreprise, dans un laboratoire de recherche (hors mathématiques) ou un organisme scientifique, d'une durée de trois semaines.

J'ai effectué ces trois semaines de stage sous la direction de Paul Fourcade à l'UFR STAPS d'Orsay, qui possède un département de recherche, le CIAMS : Complexité, Innovation et Activités Motrices et Sportives. Celui-ci est organisé autour de trois axes :

- Contrôle Moteur et Perception (CMP) ;
- Risque, Intervention, Mouvement, Equilibre (RIME) ;
- Sport, Politiques et Transformations Sociales (SPOTS).

J'ai souhaité effectuer mon stage dans cette unité car j'y ai vu l'occasion d'allier deux disciplines qui me tiennent à cœur, à savoir le sport et les sciences.

L'équipe « RIME » qui étudie entre autre l'organisation sensori-motrice du mouvement chez l'homme, notamment la formation des états posturaux statiques et dynamiques et de leurs changements. Il s'agit d'élaborer et de

1. J'ai choisi de ne pas présenter le code de cet algorithme ici, car cela nécessite d'introduire de nombreuses notations qui alourdiraient inutilement ce mémoire à mon sens.

tester des modèles dans le but de l'amélioration de la performance sportive ou encore la prévention des blessures.

Lors de ce stage, j'ai travaillé sur deux sujets totalement différents afin de pouvoir découvrir les différents aspects de la recherche que sont la modélisation à partir d'une expérience mais également l'approche expérimentale avec acquisition de données en laboratoire.

J'ai tout d'abord étudié la modélisation de coordination posturale à partir d'expériences déjà réalisées (mesure du déplacement angulaire des chevilles lors d'un mouvement d'oscillation donné). Il s'agissait de trouver une modélisation, puis de réaliser des simulations (en utilisant le logiciel Matlab) afin de voir si la modélisation s'ajustait bien avec les données initiales. Avec un modèle simple (double pendule linéaire) on ne retrouvait pas le comportement réel observé lors des expériences ; un modèle avec oscillateur non linéaire (Van der Pol) fournissait une meilleure modélisation.

Puis j'ai pris part, en tant que sujet, à la collecte de données. L'expérimentation, qui portait sur le déplacement latéral dans le but d'éviter un obstacle, a été réalisée au Centre Hospitalier Universitaire du Kremlim-Bîcetre. Chaque sujet effectuait des séries de trois mouvements différents sur une plateforme de force qui enregistrait les forces et moments exercés sur le sol lors du déplacement.

Ce stage m'a permis d'aborder différents aspects de la recherche scientifque que sont l'acquisition et la modélisation de données, ainsi que la validation d'un modèle donné. Cela m'a également donné l'occasion d'utiliser mes connaissances mathématiques et scientifiques dans un contexte différent de celui auquel je suis habituée.

# 3 Deuxième année

## 3.1 Cours

Lors de la deuxième année de Magistère, j'ai suivi les cours classiques du M1 MFA :

- Probabilités et Statistiques ;
- MAO Probabilités-Statistiques ;
- Analyse I et II ;
- Mathématiques Générales I et II ;
- Anglais ;
- Histoire des mathématiques ;

Ainsi que les cours spécifiques du Magistère :

– Compléments de théorie spectrale et d'analyse harmonique, dispensé par M. Paulin.
– Introduction aux systèmes dynamiques, par M. Breuillard.

### 3.2 TER

Il était également demandé à ce que nous réalisions un TER : Travail Encadré de Recherche. J'ai effectué ce TER avec M. Le Ny, sur le sujet de la marche aléatoire dans l'espace, avec comme référence principale [8].
Pour moi ce sujet de recherche s'inscrivait dans la continuité de cours de Probabilités de M. Cerf, qui nous avait introduit la marche aléatoire en dimension 1, et que j'avais particulièrement apprécié.
Dans ce travail, on considère un marcheur qui se trouve dans l'espace $\mathbb{Z}^d$ et se déplace selon les règles suivantes : il part de l'origine et si il est en un point $S$ au temps $n-1$, il se trouvera en l'un des $2d$ plus proches voisins de $S$ au temps $n$ avec une probabilité $\frac{1}{2d}$. On dit que le marcheur réalise une marche aléatoire en dimension $d$ et on note $S_n$ sa position au temps $n$. Le comportement du marcheur va être différent suivant la valeur de $d$. Polya a mis en évidence en 1921 une dichotomie entre le comportement planaire, où le marcheur va visiter infiniment tous les points, et spatial où il ne va les visiter qu'un nombre fini de fois.
De manière plus générale, le but de ce TER était de montrer que pour $d = 1$ ou 2 , le marcheur visite infiniment souvent tous les points alors que pour $d \geq 3$, il les visite un nombre fini de fois ; j'ai aussi étudié l'étendue de la marche aléatoire suivant la valeur de $d$, c'est-à-dire le nombre de sites visités au temps $n$.
Les notions pour cette étude sont les notions classiques de probabilités telles que convergence en probabilité et le lemme de Borel-Cantelli, ainsi que les chaînes de Markov et leurs caractéristiques : récurrence, transience, iréductiblilité.
Lors de ce TER, j'ai également réalisé des simulations à l'aide de Matlab afin d'illustrer les différentes propriétés prouvées.

## 4 Année de préparation à l'agrégation

L'année qui a suivi mon année de M1 MFA, comme beaucoup de mes collègues, j'ai fait une pause dans le cursus du Magistère afin de préparer l'agrégation de mathématiques. Ce fut également l'occasion de faire une revue générale de tous les grands domaines des mathématiques que j'avais étudiés jusque là.
Pour l'oral avec étude de document, j'avais choisi l'option Probabilités-Statistiques. J'ai obtenu l'agrégation de mathématique classée 180.

# 5 Troisième année

## 5.1 Cours

Après cette pause, c'est tout naturellement que j'ai choisi de m'inscrire au M2 Recherche Probabilités Statistiques d'Orsay. Les cours que j'ai suivis durant cette année sont :

– Concentration et sélection de modèles ;
– Séries chronologiques ;
– Statistiques en grandes dimensions ;
– Statistiques spatiales ;
– Biostatistiques ;
– Méthodes asymptotiques en statistiques ;
– Modèles pour la classification non supervisée.

Le cours spécifique pour cette dernière année de Magistère est le cours de théorie de l'information d'Elisabeth Gassiat. Ce cours, très dense, constitue une interface entre la Théorie de l'information et les statistiques. L'objectif est de transmettre des messages de la manière la plus fiable et la plus économe possible. Le message est tout d'abord codé, puis transmis, et enfin décodé. Ce cours s'intéresse à la première étape : le codage, et utilise les statistiques pour montrer les propriétés essentielles de ce premier point. Nous avons successivement abordé le codage sans perte, i.e. le codage de manière déterministe et décodable d'une suite de symboles, avec comme exemples, le codage de Shannon, le codage de Huffman et le codage de Shannon-Fano-Elias. Puis le codage en alphabet fini nous a été présenté, avec par exemple le codage de Lempel-Ziv ; et le codage en alphabet infini, avec le codage d'Élias des entiers. Il va sans dire que nous avons étudié les propriétés générales obtenues sous ces hypothèses ( alphabet fini, infini) et non pas seulement des codages particuliers. Enfin nous avons abordé la problèmatique de l'estimation statistique de l'ordre d'un modèle et son lien avec le codage universel.

Afin de valider ce cours j'ai présenté l'article [11] intitulé "Context Tree Selection : A Unifying View" de A. Garivier et F. Leonardi.

## 5.2 Stage

Lors de la recherche de mon stage, j'avais formulé un souhait : je désirais faire un stage qui me permettrait d'appliquer mes connaissances mathématiques à un problème du monde réel. Non pas que je ne sais pas apprécier la beauté d'une démonstration, mais à l'époque je n'avais pas encore décidé si je souhaitais continuer en thèse, ou bien prendre un poste de professeur de mathématiques. C'est pourquoi je voulais voir un domaine où l'on traite de vraies données, j'avais en effet pris goût au traitement de données avec

des cours plus orientés statistiques appliquées tel que les Séries chronologiques.
C'est pourquoi j'ai effectué mon stage au sein du Laboratoire de Recherche en Informatique sous l'encadrement d'Olivier Teytaud. Mon stage constituant en réalité informellement le début de ma thèse, le sujet général de ce stage et le même que ce qui est présenté dans la section suivante 6.
Au cours de ce stage, j'ai étudié et fait la preuve de convergence d'un algorithme d'optimisation bruitée utilisant l'information du second ordre (i.e. gradient et Hessienne) ; pour tous ces termes, je vous invite à consulter la section 6.5.1.

# 6 Présentation de mon sujet de thèse

## 6.1 Introduction

Le sujet de ma thèse s'intitule "Systèmes électriques de grande taille : Statistique et Optimisation", effectuée sous la direction d'Olivier Teytaud, au sein de l'équipe TAO, LRI, INRIA, CNRS UMR 8623 ; et est financée par le projet POST.

L'optimisation de systèmes électriques de grande taille pose de beaux problèmes d'optimisation, dus entre autre à :

i des incertitudes stochastiques ;
ii de grandes dimensions : typiquement échelle Europe + Afrique du Nord ;
iii de multiples niveaux d'anticipation :
  - Problème d'investissement à haut niveau : décider par exemple, combien d'éoliennes implanter et où dans les 30 prochaines années ;
  - Problème de gestion stratégique à bas niveau : quelle stratégie utiliser sur une centrale hydroélectrique à l'échelle un mois ?

iv la non linéarité ;
v la dispersion géographique des moyens de production ;

ce qui ne constitue pas une liste exhaustive.
Selon l'expertise de [12] la part d'électricité issue d'énergies renouvelables dans la production mondiale en 2050 pourrait être de 10% pour les scénarios les plus pessimistes, ou bien de 50% pour les scénarios les plus hauts. Cette utilisation accrue d'énergies renouvelables complique le problème, en effet, la variabilité de la production nécessite une stabilisation du système par stockage, transport, moyennage sur de grandes zones géographiques. De plus, le concept d'un seul réseau de production et de distribution d'énergie à l'échelle mondiale est développé dans [6], où il est dit que ce modèle pourrait être réalisable du point de vue technique et compétitif au niveau économique. Ce réseau, en particulier, serait à même de capter les sources d'énergies isolées (centrales solaires photovoltaïques au Sahara ou parcs éoliens au Groenland,

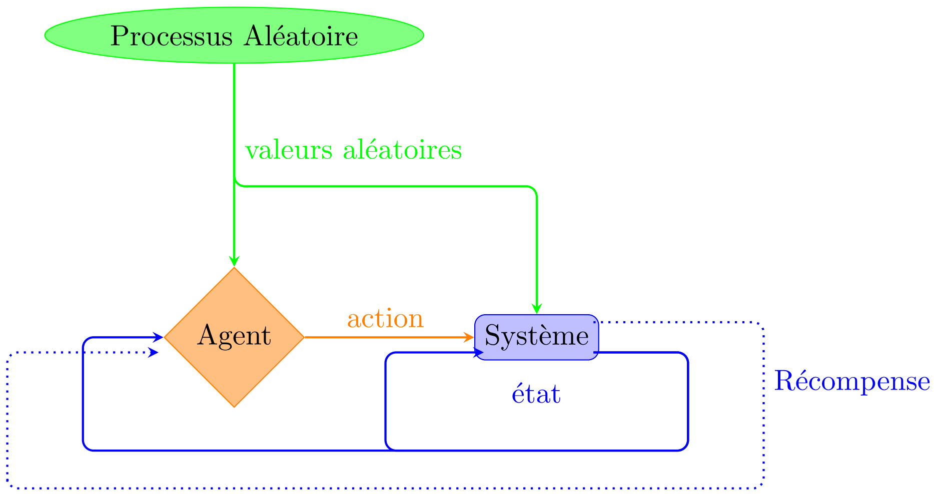


Figure 1 – Prise de Décision Séquentielle

par exemple), et de transporter cette électricité sur de longues distances.

Tout d'abord, la section 6.2 décrit le modèle utilisé, puis la section 6.3 montre un aperçu des différentes techniques classiques de résolution de ce problème, ainsi que leurs limites dans le cas qui nous préoccupe. La section 6.4, introduit de nouvelles idées afin d'apporter une réponse adéquate à ces questions. Enfin la section 6.5 donne une idée des outils mathématiques susceptibles d'être utilisés pour étudier le problème.

## 6.2 Modélisation du problème : Prise de Décision Séquentielle

La gestion d'un système électrique est un Problème de Prise de Décision Séquentielle, également désigné sous l'appellation programme dynamique, problème de prise de décision multiétagée ou encore prise de décision sous incertitudes. Nous appelerons ce problème le PPDS pour plus de concision. La description du problème est empruntée aux très bons [3, 7, 14].

Dans le PPDS (cf figure 1), un agent ( également appelé contrôleur) interagit avec un système (aussi dénommé environnement, ou processus) durant un certain nombre de pas de temps successifs. À chaque pas de temps :

1. le système communique son état courant à l'agent ;
2. l'agent, en retour, envoie des commandes ou actions au système ;
3. le système change d'état conformément à ces commandes (suivant une fonction de transition) ;
4. le système renvoie la récompense correspondante.

Des variables aléatoires peuvent intervenir dans le processus, comme illustré sur la figure 1, à plusieurs niveaux. L'agent peut obtenir une information sur l'état courant altérée (information bruitée) due par exemple à des erreurs de mesure ; un processus aléatoire peut contribuer au changement d'état du système, penser à l'impact de la météo sur la consommation électrique. L'objectif du contrôleur est de choisir des actions de sorte que la récompense à long terme soit aussi grande que possible. Les 'pas de temps', 'états', 'actions', 'processus aléatoire', 'fonction de transition' et 'récompense' sont les différents éléments qui composent le PPDS. Ces concepts sont brièvement décrits ci-dessus, en soulignant les spécificités du problème de gestion d'un système électrique.

### 6.2.1 Pas de temps

**Définition 2.** *Les pas de temps sont les instants auxquels une décision doit être prise, i.e. auxquels le contrôleur doit choisir une action.*

En général, l'ensemble des instants de prise de décision peut être discret ou continu, ici on considère un ensemble discret : $t \in \{1, \ldots, N\}$, d'où l'appellation 'pas de temps' ; discret ne signifie pas forcément durées égales entre les pas de temps. Lorsque c'est pertinent, on peut indexer les variables avec leurs pas de temps respectifs, en notant $X_t$ la variable $X$ au temps $t$.

### 6.2.2 Variables d'état

**Définition 3.** *La variable d'état est la fonction de dimension minimum du passé du système nécessaire et suffisante pour déterminer, via les fonctions adéquates, la prochaine action, le nouvel état et la récompense correspondante.*

Ainsi, une variable d'état peut inclure des informations concernant l'état physique du système ou encore des informations sur les pas de temps précédents (en cas de non stationnarité) si nécessaire.
On note $\mathcal{X}$ l'espace d'états, qui peut être fini, infini dénombrable ou indénombrable ; si besoin est, un espace d'état continu peut être discrétisé. La nature de $\mathcal{X}$ est déterminée à la fois par le problème même et par les hypothèses formulées sur le modèle. Une variable d'état est notée $x \in \mathcal{X}$, ou éventuellement $x_t$ si elle dépend du pas de temps $t$.

### 6.2.3 Variables d'action

**Définition 4.** *Les actions sont les commandes données par l'agent au système, choisies de sorte que la récompense totale est maximale. Une action réalisable ou légale est une action qu'il est possible de choisir étant donné un état $x$, à un temps $t$.*

L'espace d'actions est noté $\mathcal{A}$, et une action $a \in \mathcal{A}$, éventuellement $a_t$ si elle dépend du pas de temps $t$. Comme pour l'espace d'états, $\mathcal{A}$ peut être fini, infini dénombrable ou indénombrable, et discrétisé si nécessaire. Parfois, l'agent doit connaître les décisions passées pour choisir l'action suivante, dans ce cas, les actions sont également inclues dans les variables d'état.

### 6.2.4 Processus aléatoire

On note $\omega_t$ la valeur du processus aléatoire au temps $t$, et $\Omega_t$ l'ensemble des réalisations possibles à $t$ (éventuellement sans l'indexation en $t$ en cas de stationnarité). Il faut remarquer qu'au pas de temps $t$, le contrôleur connaît $\omega_t$ mais $\omega_{t+1}$ est stochastique. Comme précédemment, $\Omega_t$ peut être fini, infini dénombrable ou indénombrable. Le processus aléatoire suit une distribution inconnue, dépendant potentiellement du temps, des états et des actions. Si $\omega_t$ dépend du passé du processus, alors il appartient aussi aux variables d'état : $\omega_t \in x_t$.
La difficulté majeure dans la prise de décision séquentielle sous incertitudes est de savoir comment prendre en compte cet aspect stochastique. Certaines méthodes exigent d'avoir à disposition un modèle permettant d'engendrer une nouvelle réalisation du processus à chaque fois que c'est nécessaire : on travaille dans ce cadre, en ajustant par exemple un modèle sur un ensemble de réalisations données. D'autres méthodes recherchent une solution optimale en considérant au choix :

1. un unique scénario donné ;
2. la moyenne du processus aléatoire ;
3. les 5% des pires cas ;
4. un processus markovien.

Les points 1, 2 and 3 permettent une optimisation bien plus simple, cependant, passer outre le caractère aléatoire du problème peut conduire à des solutions terriblement mauvaises, par exemple, si un épisode de forte chaleur a lieu en été, la consommation électrique sera très différente d'un scénario donné sans période caniculaire. Le point 4 présente également des faiblesses lorsque le processus sous-jacent n'est pas markovien. C'est pourquoi il est primordial de faire un bon compromis entre :

– des hypothèses trop simples, menant à un problème d'optimisation facile mais apportant potentiellement de mauvaises solutions en conditions réelles ;
– des hypothèses plus contraignantes, donnant lieu à un problème d'optimisation difficile.

### 6.2.5 Fonction de transition

**Définition 5.** *La fonction de transition modélise la dynamique du problème. Étant donné un pas de temps $t$, un état $x_t$, une action $a_t$ et un processus aléatoire $(\omega_t)_{1\leq t\leq N}$, la fonction de transition $f$ fournit le nouvel état $x_{t+1}$ du système : $x_{t+1} = f(x_t, a_t, \omega_{t+1})$.*

Dans un contexte stochastique, il faut noter que $(x_t)_{1\leq t\leq N}$ suit un processus aléatoire inconnu.

### 6.2.6 Fonction de récompense

**Définition 6.** *La fonction de récompense $g$ permet de quantifier les préférences de l'optimiseur. Étant donné un pas de temps $t$, un état $x_t$, une action $a_t$ et un processus aléatoire $(\omega_t)_{1\leq t\leq N}$, on note $r_t = g(x_t, a_t, \omega_{t+1})$.*

Le contrôleur a alors pour but de maximiser la récompense totale :

$$R_N = \sum_{t=1}^{N} \gamma^t r_t, \text{ où } 0 < \gamma < 1. \tag{1}$$

Lorsque $R_N$ est aléatoire, l'optimiseur peut décider de maximiser au choix : l'espérance de la récompense, la récompense sur les 5% des pires cas ou bien encore la récompense sur un scénario donné. On choisit ici de maximiser l'espérance de la récompense.

## 6.3 Méthodes de résolution du PPDS

On introduit deux notions utiles pour résoudre le PPDS :

- la notion de politique, qui définit la façon dont le contrôleur choisit les actions ;
- la fonction valeur $x \mapsto V^\pi(x)$, qui est définie comme la récompense obtenue par un agent commençant à l'état $x$ et appliquant la politique $\pi$ jusqu'à atteindre un état final.

### 6.3.1 Politiques

**Définition 7.** *Une politique est une règle, ou une fonction, éventuellement stochastique, qui détermine une décision étant donnée l'information disponible à l'état $x_t$. Soient $\pi$ une politique, $t$ un pas de temps, $x_t \in \mathcal{X}$ un état, l'action sélectionnée est $a_t \in \mathcal{A}$ avec probabilité $\pi(x_t, a_t)$. Par commodité, on note $\pi(x_t)$ la valeur aléatoire qui vaut $a_t$ avec probabilité $\pi(x_t, a_t)$.*

Pour donner une idée, voici quelques politiques classiques :

- politique aveugle : une politique choisie au début, qui ne prend pas en compte l'état courant du système. Formellement $t \in x_t$ (i.e le temps fait partie de la variable d'état) et $\pi(x_t) = \pi(t)$ (i.e. $\pi$ ne dépend que de $t$) ;

– réseau de neurones : typiquement $\pi_{\theta_1,\theta_2,\theta_3,\theta_4}(x) = \theta_1 + \theta_4 \times sigmoide(\theta_2.x + \theta_3)$, où $\theta_1, \theta_2, \theta_3, \theta_4$ sont des réels, et $x$ est l'état courant du système ;
– polynomiale : $\pi_{\theta_1,\theta_2}(x) = \theta_1 + \theta_2.x$ .

### 6.3.2 Fonction Valeur

**Définition 8.** *$V^\pi : \mathcal{X} \to \mathbb{R}$ est définie comme suit :*

$$V^\pi(x_{t_0}) = \mathbb{E}_\pi \left[ \sum_{t_0 \le t \le N} \gamma^t r_t | x_{t_0} \right] \tag{2}$$

*où $\mathbb{E}_\pi[X]$ est l'espérance de $X$ sous l'hypothèse qu'à chaque pas de temps, l'agent applique la politique $\pi$.*
*Ainsi, si $\Pi$ est l'ensemble des politiques possibles $\pi$, alors étant donné un état initial $x_0$, une politique optimale $\pi^*$ vérifie :*

$$V^{\pi^*}(x_{t_0}) = \sup_{\pi \in \Pi} V^\pi(x_{t_0}) = V^*(x_{t_0}) \tag{3}$$

$V^{\pi^*}(x_{t_0})$ est appelée valeur optimale de $x_{t_0}$, c'est la plus grande récompense moyenne qu'on peut obtenir en commençant à l'état $x_{t_0}$.

### 6.3.3 Recherche Directe de Politique

Une première méthode à laquelle on peut songer pour résoudre le PPSD est la méthode de Recherche Directe de Politique.

On suppose que la politique $\pi$ est une fonction paramétrique, dépendant d'un paramètre $\theta$. On note cette politique $\pi_\theta$ et $V_\theta$ la fonction valeur associée. Soit $x_0$ un état initial, un algorithme de type boîte noire (cf algorithme 2, où on utilise les notations précédemment introduites) nous donne une valeur bruitée de $V_\theta(x_0)$, on optimise ainsi $\theta \mapsto V_\theta(x_0)$ à l'aide d'un algorithme d'optimisation bruitée (cf section 6.5.1).

**Algorithm 2** Algorithme de type boîte noire.

```
Entrées : paramètre θ, état initial x_0
r ← 0
s ← x_0
Tantque s ≠ état final
ω ← aléa
a = π_θ(s)
r+ = g(s, a, ω)
s = f(s, a, ω)
FinTantque
Retourner r
```

L'avantage de cette technique est qu'elle impose peu de contraintes : (i) sur l'action $a$ car $\pi_\theta$ n'est pas forcément linéaire et (ii) sur le processus aléatoire $(\omega_t)_{t\geq 1}$ qui n'est pas forcément markovien. D'un autre côté, dans des problèmes spécifiques, l'action doit satisfaire des hypothèses que la politique ne peut remplir (par exemple, dans le cas d'une centrale nucléaire, la production d'énergie ne peut être en-dessous d'un niveau donné). De plus, si l'espace d'actions est très grand, ce qui est le cas dans notre étude, alors le paramètre $\theta$ sera de grande dimension, rendant l'optimisation d'autant plus compliquée.

### 6.3.4 Équations de Bellman

L'utilisation des équations de Bellman constitue une autre méthode pour résoudre le PPDS via la fonction de valeur, lorsque les espaces d'états et d'actions sont finis.

**Définition 9** (Équations de Bellman pour la fonction de valeur)**.** *La fonction de valeur de la politique $\pi$ satisfait l'équation suivante :*

$$V^\pi(x) = \mathbb{E}_\omega \left\{ g(x, \pi(x), \omega) + \gamma V^\pi(f(x, \pi(x), \omega)) \right\} \tag{4}$$

*Et la fonction de valeur optimale vérifie l'équation suivante :*

$$V^*(x) = \sup_{a\in\mathcal{A}} \mathbb{E}_\omega \left\{ g(x, a, \omega) + \gamma V^*(f(x, a, \omega)) \right\} \tag{5}$$

**Idée générale :** Algorithme à rebours.
Pour tout état initial $x_{t_0} \in \mathcal{X}$, la récompense optimal $V^{\pi^*}(x_{t_0}) = V(x_{t_0})$ est donnée par la dernière étape de cet algorithme à rebours :

$$V(x_N) = g(x_N) \tag{6}$$

$$V(x_t) = \max_{a_t\in\mathcal{A}(x_t)} \mathbb{E}_{\omega_{t+1}} \left\{ g(x_t, a_t, \omega_{t+1}) + V(f(x_t, a_t, \omega_{t+1})) \right\} \tag{7}$$

pour $t = 0, \ldots, N-1$, où la moyenne est calculée par rapport à la distribution de probabilité du processus $\omega_{t+1}$, dependant de $x_t$ et $a_t$. De plus, si $a_t^* = \pi_t^*(x_t)$ maximise l'expression à droite de l'égalité dans l'équation (7) pour tout $x_t \in \mathcal{X}$ et $t \in \{1, \ldots, N\}$, alors $\pi^* = \pi_0^*, \ldots, \pi_{N-1}^*$ est optimal.
Le coût de calcul de cet algorithme est en $O\left(|\mathcal{A}||\mathcal{X}|^2\right)$, où $|E|$ désigne le cardinal de l'ensemble $E$. Lorsque le cardinal de $\mathcal{X}$ est grand, cette technique n'est pas appropriée, ce qui est le cas dans notre problème. Un autre inconvénient vient du fait que dans cette méthode, le processus aléatoire doit être markovien, ce qui n'est pas nécessairement vrai dans notre cadre.

## 6.4 Recherche Directe de Valeur

Dans la section précédente, nous avons présenté deux méthodes pour résoudre le problème de prise de décision séquentielle sous incertitudes. Ces deux méthodes, Recherche Directe de Politique et Équations de Bellman, ont toutes les deux des avantages et des inconvénients. C'est pourquoi nous proposons une nouvelle méthode à développer, qui tire partie de ces solutions.
L'idée est de garder la forme des équations de Bellman (4), mais on substitue à la fonction de valeur définie dans (2), une fonction paramétrique de la forme de celle utilisée dans la Recherche Directe de Politique.

**Définition 10** (Recherche Directe de Valeur)**.** *La Recherche Directe de Valeur consiste à trouver la valeur :*

$$\arg\max_{a\in\mathcal{A}} \mathbb{E}_\omega \left\{ g(x_t, a, \omega) + \Pi_\theta(x_t, x_{t+1}, t+1) \right\} \tag{8}$$

*où :*

1. $x \mapsto \Pi_\theta(x, x', t)$ *et* $t \mapsto \Pi_\theta(x, x', t)$ *peuvent ne pas être linéaires ;*
2. $x' \mapsto \Pi_\theta(x, x', t)$ *est une fonction linéaire convexe par morceaux.*

Dans un tel cadre, grâce à 1, des contrôles complexes peuvent être encodés : on bénéficie du caractère générique de la Recherche Directe de Politique ; et 2 permet d'effectuer une optimisation linéaire, ce qui autorise un grand espace d'actions. Enfin, le processus $(\omega_t)$ n'est pas nécessairement markovien.

## 6.5 Développements envisagés : quelques pistes

Pour conclure, on donne un aperçu des principaux éléments algorithmiques et mathématiques susceptibles d'apporter des réponses au problème. On présente tout d'abord des algorithmes adaptés, puis on évoque quels propriétés classiques en probabilité.

### 6.5.1 Algorithmes d'optimisation

**Définition 11** (Fonction bruitée)**.** *Soit* $\mathcal{D} \in \mathbb{R}^d$*, une fonction bruitée* $f$ *est un processus aléatoire :* $(x, \omega) \mapsto f(x, \omega)$*, où* $x \in \mathcal{D}$ *et* $\omega$ *est une variable aléatoire. À chaque appel à la boîte noire, l'utilisateur fournit* $x$ *uniquement, et c'est une nouvelle réalisation indépendante de* $\omega$ *qui est appliquée.*

L'optimisation bruitée est l'optimisation d'une fonction objectif bruitée : il faut trouver $x$ maximisant $x \mapsto \mathbb{E}[f(x, \omega)]$. On travaille dans le cadre de boîte noire, c'est-à-dire qu'on n'utilise aucune propriété interne de $f$, on a accès uniquement à des réalisations de la variable aléatoire $f(x, \omega)$ lorsque l'on fournit $x \in \mathcal{D}$ à la boîte noire.

On propose et on étudie différents types d'algorithmes d'optimisation bruitée :

- des algorithmes utilisant l'information d'ordre zéro (c'est-à-dire des méthodes n'utilisant ni gradient ni Hessienne) , tels que ceux désignés sous l'appellation Stratégie d'Évolution, se référer à [4] par exemple ;
- des algorithmes utilisant l'information du premier ordre, i.e. le gradient (estimé par différences finies par exemple) comme dans [9, 15] ;
- des algorithmes utilisant l'information du second ordre, i.e. le gradient et la Hessienne comme dans [10].

Le but étant d'améliorer leur taux de convergence en adaptant les paramètres par exemple ; on peut également développer de nouveaux algorithmes robustes et efficaces en combinant des algorithmes déjà existants ou en adaptant des algorithmes efficaces dans le cas non bruité.
Un autre objectif est de montrer des propriétés mathématiques de convergence ou de divergence de ces algorithmes, dépendant des hypothèse formulées sur la fonction objectif.

### 6.5.2 Bagage probabiliste

Les outils utiles pour travailler sur ces sujets sont, à ma connaissance, et ce n'est pas une liste exhaustive, les outils classiques de probabilité que sont les inégalités de Markov et de Tchebychev ainsi que des inégalités de concentration, voir par exemple [5].

**Propriété 1** (Inégalité de Markov)**.** *Soit $X$ une variable aléatoire positive, alors pour tout réel $\epsilon > 0$*

$$P(X \geq \epsilon) \leq \frac{\mathbb{E}(X)}{\epsilon}$$

**Propriété 2** (Inégalité de Tchebychev)**.** *Soit $X$ une variable aléatoire telle que $\mathbb{E}(X) = \mu < \infty$ et $0 < Var(X) = \sigma^2 < \infty$, alors pour tout réel $\epsilon > 0$,*

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Les propriétés sur les chaînes de Markov telles que pérodicité, réductibilité, récurrence, stabilité, ergodicité, etc ... voir [2, 13] à ce sujet, peuvent aussi être utiles pour montrer la convergence d'un algorithme.

**Définition 12** (Chaîne de Markov)**.** *Une chaîne de Markov est une séquence de variables aléatoires $X_1, X_2, \ldots$ telles que pour tout entier $n$, pour toute sequence $(x_1, \ldots, x_n, x)$,*

$$P(X_{n+1} = x | X_1 = x_1, \ldots, X_n = x_n) = P(X_{n=1} = x | X_n = x_n)$$

Les propriétés issues de la théorie de la mesure telle que le théorème de convergence dominée sont également utiles ici.

# 7 Conclusion

Cette rapide rétrospective sur ces trois années d'étude tend à illustrer quelle type de formation propose le Magistère de mathématiques d'Orsay ; formation qui correspondait bien à mes attentes. Le Magistère a dispensé une bonne formation générale, donnant aux étudiants l'expérience nécessaire pour continuer dans l'enseignement ou dans la recherche, via :

- un enseignement large et approfondi de différents grands domaines des mathématiques permettant un socle de connaissances solides, étendues et diversifiées ;
- des stages et travaux de recherche favorisant l'expression orale et la présentation claire d'un domaine de recherche ;
- un maximum d'autonomie, en nous poussant à trouver un projet de stage, de recherche ;
- des contacts directs avec des chercheurs, ce qui facilite l'entrée en thèse.

Durant ces trois années d'études, j'ai acquis quantité de savoirs qui me serviront tous un jour ou l'autre, soit directement pendant ma thèse, tels que les outils de probabilités, statistiques, analyse, algèbre, informatique et algorithmiques ; soit indirectement, de par l'ouverture d'esprit que cela m'aura apporté.

# Références


[1] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. *Annals of Mathematics,year=2004,page=781-793,volume=160.*

[2] Anne Auger. Convergence results for (1-$\lambda$)-SA-ES using the theory of $\varphi$-irreductible Markovs chains. *Theoretical Computer Science,year=2005,page=35-69,volume=334.*

[3] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume I and II. Athena Scientific, 1995.

[4] H-G. Beyer. *The Theory of Evolution Strategies.* Natural Computing Series. Springer-Verlag, Netherlands, 2001.

[5] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities : A Nonasymptotic Theory of independence.* Oxford University Press, 2013.

[6] Spyros Chatzivasileiadis, Damien Ernst, and Göran Andersson. The global grid. *Renewable Energy*, 87, september 2013.

[7] Adrien Couëtoux. *Monte Carlo Tree Search for Continuous and Stochastic Sequential Decision Making Problems.* PhD thesis, 2013.

[8] A. Dvoretzky and P. Erdös. *Some problems on random walk in space.* Berkeley Symposium on Mathematical Statistics and Probability, 1951.

[9] Vaclav Fabian. Stochastic Approximation of Minima with Improved Asymptotic Speed. *Annals of Mathematical statistics*, 38 :191–200, 1967.

[10] Vaclav Fabian. *Stochastic Approximation.* SLP. Department of Statistics and Probability, Michigan State University, 1971.

[11] A. Garivier and F. Leonardi. Context tree selection : A Unifying View. *Stochastic Processes and their Applications,year=2011.*

[12] Eric Martinot, Carmen Dienst, and Weiliang Liu. Renewable energy futures : Targets, scenarios, and pathways. *Annual Review of Environment and Resources*, 32 :205–239, 2007.

[13] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, New-York, 1993.

[14] Warren B. Powell. *Approximate Dynamic Programming : Solving the curse of dimensionality.* Wiley series in Probability and Statistics. Wiley-Interscience, 2007.

[15] Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. *CoRR*, abs/1209.2388, 2012.