

**EA d'initiation à la recherche
du PA de Mathématiques Appliquées**

APM_51111_EP – APM_52112_EP – APM_50113_EP

3 avril 2025

Table des matières

Introduction - EA d'initiation à la recherche	3
Choix et attribution des projets	3
Déroulement du projet, encadrement	5
Objectifs, rapport et soutenance, notation	5
1 Data science : son & image, réseaux, apprentissage	7
1.1 Traitement des images	7
1.2 Machine learning : Théorie et applications	8
1.3 Machine learning : Étude de cas	8
2 Recherche Opérationnelle	9
2.1 Applications de la recherche opérationnelle	10
2.2 Fondements mathématiques de la recherche opérationnelle	10
3 Modèles aléatoires pour la physique, l'informatique et la biologie	12
3.1 Modèles aléatoires pour la biologie	13
3.2 Modèles aléatoires pour la physique ou l'informatique	13
4 Modélisation et ingénierie financière	14
4.1 Évaluation et gestion des risques	14
4.2 Interactions stratégiques et risque systémique	15
4.3 Statistique haute fréquence et microstructure des marchés	15
5 Simulation Aléatoire	16
5.1 Processus aléatoires	16
5.2 Quantification d'incertitude	17
6 EDP, analyse numérique et applications en physique, mécanique et biologie	18
6.1 Sous-thématique : Modélisation et simulation	18
6.2 Sous-thématique : Analyse numérique et calcul scientifique	18
6.3 Sous-thématique : Optimisation déterministe	19

Introduction - EA d'initiation à la recherche

Objectifs

Le but de ces enseignements est de fournir une initiation à la recherche et développement en mathématiques appliquées, à travers la réalisation d'un projet. Le projet consiste en l'étude d'un problème, motivé par les applications ou des questions de nature mathématique, allant de la modélisation à l'implémentation numérique et à l'analyse critique des résultats.

Ce projet est effectué en binôme ou en trinôme, et constitue un véritable travail d'équipe.

Choix et attribution des projets

Une fois que votre inscription dans l'une des EA d'initiation à la recherche aura été validée, il vous sera demandé de constituer des groupes (binômes ou trinômes) et de faire des vœux, afin de déterminer le domaine dans lequel s'effectuera votre projet. Les thématiques disponibles sont listées ci-dessous (en *italique* le référent pour chaque thématique). Elles sont divisées en sous-thématiques, décrites plus en détail à la fin de ce document, et vous trouverez de courtes vidéos de présentation sur Moodle. **La liste de sous-thématiques ci-dessous correspond à celle proposée en 2024-2025, et pourrait être légèrement modifiée l'année prochaine.**

1. **Data science** (*El Mahdi El Mhamdi*)
 - 1.1 Traitement des images
 - 1.2 Machine learning : Théorie et application
 - 1.3 Machine learning : Étude de cas
2. **Recherche opérationnelle** (*Xavier Allamigeon, Stéphane Gaubert*)
 - 2.1 Applications de la recherche opérationnelle
 - 2.2 Fondements mathématiques de la recherche opérationnelle
3. **Modèles aléatoires pour la biologie, la physique et l'informatique** (*Lucas Gerin*)
 - 3.1 Modèles aléatoires pour la biologie
 - 3.2 Modèles aléatoires pour la physique ou l'informatique
4. **Modélisation et ingénierie financière** (*Stefano De Marco*)
 - 4.1 Évaluation et gestion des risques
 - 4.2 Interactions stratégiques et risque systémique
 - 4.3 Statistique haute fréquence et microstructure des marchés
5. **Simulation Aléatoire** (*Clément Rey*)

5.1 Processus aléatoires

5.2 Quantification d'incertitude

6. **Équations aux dérivées partielles, analyse numérique et applications en physique, mécanique et biologie** (*Teddy Pichard*)

6.1 Modélisation et simulation

6.2 Analyse numérique et calcul scientifique

6.3 Optimisation déterministe

▷ **Enseignants référents**

Pour chacune des thématiques, les enseignants référents sont disponibles pour répondre à vos questions. Leurs coordonnées sont indiquées dans la suite de ce document. N'hésitez pas à prendre contact avec eux pour tout souci pendant la durée du projet (problème d'organisation, problème scientifique...).

▷ **Procédure de vœux**

Chaque groupe devra effectuer des vœux, en remplissant un formulaire en ligne. Dans ce formulaire, vous devrez indiquer **5 choix de sous-thématiques, classés par ordre de préférence, en indiquant le numéro des sous-thématiques (1.3, 6.1, 2.2...)**. Il est obligatoire que vos vœux portent sur au moins **3 thématiques différentes**. Ce formulaire sera disponible sur Moodle une fois que les inscriptions pour la 3A seront clôturées et validées, vous serez contactés à ce moment-là. Il devra être rempli avant

date limite en juin à déterminer

Tous les choix envoyés à temps et respectant les règles ci-dessus seront traités de façon égalitaire.

▷ **Formation des groupes**

Vous êtes libres de former des groupes (binômes ou trinômes) comme vous l'entendez. Ces groupes devront être renseignés dans le formulaire en ligne, et **une seule soumission par groupe suffit**. Si vous ne parvenez pas à former un groupe, vous pourrez remplir le formulaire seul, mais vous serez alors automatiquement associé à une autre personne seule ou à un groupe existant.

▷ **Articulation entre projets**

Si vous êtes inscrits à APM_51111_EP et APM_52112_EP, il ne sera pas possible d'effectuer deux projets dans la même thématique. Si vous souhaitez faire un seul projet plus long sur le même sujet, il convient de s'inscrire plutôt à APM_50113_EP.

▷ Attribution des encadrants

Après la période des vœux, un encadrant sera attribué à chaque groupe, en essayant de respecter au mieux les choix des sous-thématiques envoyés à temps.

Si vous vous inscrivez ou envoyez vos vœux hors délai, ou que ceux-ci ne respectent pas les contraintes décrites ci-dessus, il sera plus compliqué d'obtenir un encadrant dans une des sous-thématiques de votre choix. Même en cas de retard/changement d'EA, il sera impossible de rejoindre les EA APM_51111_EP et APM_50113_EP après le 1er septembre, et l'EA APM_52112_EP après le 14 décembre.

Déroulement du projet, encadrement

▷ Démarrage du projet

Dès que votre encadrant vous aura été attribué, **vous devrez prendre contact avec lui**, afin de fixer la date d'un premier rendez-vous, et de définir précisément le sujet sur lequel vous allez travailler. Cet encadrant vous précisera les modalités de travail spécifiques au sujet, et supervisera votre travail tout au long du projet.

▷ Déroulement du projet

Ce projet constitue un travail personnel dont l'intérêt et la richesse dépendront principalement de votre investissement. L'enseignant qui vous encadre vous guidera dans votre démarche. Surtout n'hésitez pas à le contacter.

C'est à vous de le solliciter (et non l'inverse).

Le créneau du mardi après-midi est consacré au travail sur vos projets. Les rencontres avec votre encadrant pourront avoir lieu aussi à d'autres moments, en fonction de vos emplois du temps respectifs.

Notation, rapport et soutenance

▷ Rapport et soutenance

- Pour les projets sur une période (APM_51111_EP et APM_52112_EP), l'évaluation aura lieu à la fin de la période concernée.
- Pour les projets longs (APM_50113_EP), une première évaluation de mi-parcours aura lieu en fin de période 1, et une évaluation finale aura lieu en fin de période 2.

Chacune de ces évaluations sera basée sur la remise d'un rapport écrit, et sur une soutenance orale.

Les rapports sont à envoyer en format pdf à tous les membres du jury de soutenance (cc l'enseignant responsable de la thématique).

Les soutenances sont typiquement composées d'un exposé de 30 minutes environ (partagées équitablement entre les élèves) suivi de 10 minutes de questions. Les dates exactes et les horaires des soutenances vous seront transmis par le département en temps voulu.

Indications. Dans le rapport comme lors de la soutenance, une bonne pratique est d'agir comme si le jury ne connaissait pas le problème, et donc l'introduire en montrant son importance, expliquer votre approche ainsi que le cadre théorique, et montrer vos résultats ; vous donnerez enfin vos conclusions en faisant un bilan de votre travail.

Le rapport, de préférence écrit en $\text{\LaTeX}$, devra comprendre une bibliographie des ouvrages et articles étudiés. Vous êtes encouragés à chercher de la documentation sur votre sujet.

▷ **Évaluation**

Elle tiendra compte de la qualité du contenu du rapport et de la soutenance, ainsi que des interactions avec l'enseignant. Elle prendra également en compte le sens critique sur les résultats obtenus et la précision de la bibliographie. Une seule note est généralement attribuée à l'ensemble du groupe.

Tous les sujets proposés demanderont d'effectuer des simulations/calculs numériques. Les outils informatiques seront mis en œuvre comme un moyen d'expérimentation et de compréhension, et non comme une fin en soi.

Bon travail à tous,

les équipes pédagogiques APM_51111_EP, APM_52112_EP, APM_50113_EP.

1 Data science : son & image, réseaux, apprentissage

Enseignant référent : El Mahdi El Mhamdi el-mahdi.el-mhamdi@polytechnique.edu

Les sujets proposés dans la thématique Data Science visent à donner une expérience des méthodes de Machine Learning utilisées dans le traitement automatique des données, que ce soit d'un point de vue théorique ou appliqué. Les sujets sont ainsi répartis dans les 3 sous-thématiques suivantes : traitement des images, Machine learning : théorie et applications, Machine learning : Étude de cas.

La première catégorie est, comme son nom l'indique, un champs d'applications bien spécifique : les méthodes utilisées pour des données comprenant des images sont le plus souvent basées sur des réseaux de neurones, mais d'autres approches plus classiques peuvent être envisagées. La deuxième thématique s'adresse à tous les étudiants souhaitant aborder un problème sous l'aspect théorique/méthodologique avec une application simple à un problème réel. La dernière sous-thématique correspond à une approche au plus près des données comportant peu de théorie et centrée sur la résolution concrète d'un véritable problème impliquant des données réelles.

1.1 Traitement des images

Présentation de la sous-thématique. Au même titre que les textes, les images sont des données requérant un traitement spécifique, du fait des fortes corrélations présentes entre pixels voisins dans une image. Les méthodes actuelles les plus performantes pour effectuer de la classification d'images, ou de la segmentation, utilisent des réseaux de neurones profonds. Le but de cette sous-thématique est d'étudier des méthodes de réseaux de neurones ainsi que des méthodes plus classiques, afin notamment de segmenter automatiquement des images ou de détecter des biais dans les algorithmes de classification préexistants.

Keywords : Modelling, Image segmentation, Bias, Deep Learning, Image segmentation.

- [1] Leo Grady, *Random walks for image segmentation*, IEEE transactions on pattern analysis and machine intelligence **28** (2006), no. 11, 1768–1783.
- [2] Alexandra Chouldechova, *Fair prediction with disparate impact : A study of bias in recidivism prediction instruments*, Big data **5** (2017), no. 2, 153–163.
- [3] Philippe and del Barrio Besse Eustasio and Gordaliza, *A survey of bias in Machine Learning through the prism of Statistical Parity for the Adult Data Set*, arXiv preprint arXiv :2003.14263 (2020).
- [4] Joy and Gebru Buolamwini Timnit, *Gender shades : Intersectional accuracy disparities in commercial gender classification*, Conference on fairness, accountability and transparency, 2018, pp. 77–91.

1.2 Machine learning : Théorie et applications

Présentation de la sous-thématique. Tous les sujets proposés dans cette thématique sont centrés sur l'étude d'un algorithme spécifique (par exemple des algorithmes d'optimisation, des réseaux de neurones...) ou d'un problème générique (détection de communauté dans un graphe...) nécessitant l'utilisation d'un algorithme spécifique. La majorité du projet de recherche consiste à comprendre les différents algorithmes proposés d'un point de vue méthodologique et théorique, puis à appliquer cet algorithme sur un jeu de données réelles ou simulées. Les possibilités sont variées, le point commun de tous ces sujets étant l'aspect théorique/méthodologique.

Keywords : Deep Learning, Transfer Learning, Optimal Transport, Structured Data, graphs, optimization, deep reinforcement learning.

- [1] Richard Sutton and Andrew G Barto, *Reinforcement learning : An introduction*, MIT press, 2018.
- [2] Emmanuel Abbe, *Community detection and stochastic block models : recent developments*, ArXiv e-prints (2017), available at [1703.10146](https://arxiv.org/abs/1703.10146).
- [3] S. Gadat, *Stochastic Optimization algorithms*, M2 Lecture Notes.
- [4] Y. Nesterov, *Algorithms for reinforcement learning*, Efficiency of coordinate descent methods on huge-scale optimization problems **22** (2010), 341–362.
- [5] Subhashis and Ghosh Ghosal Jayanta K. and van der Vaart, *Convergence rates of posterior distributions*, Annals of Statistics **28** (2000), no. 2, 500–531.
- [6] L. and Györfi Devroye L. and Lugosi, *A probabilistic theory of pattern recognition*, Applications of Mathematics (New York), vol. 31, Springer-Verlag, New York, 1996.

1.3 Machine learning : Étude de cas

Présentation de la sous-thématique. Chaque sujet proposé dans cette sous-thématique vise à étudier un jeu de données précis. Vous pouvez voir cette sous-thématique comme un challenge kaggle : aucun développement théorique n'est attendu. En revanche, il faudra proposer différentes méthodes pour résoudre le problème posé et comparer empiriquement ces méthodes de façon précise. Les données et problèmes proposés ici seront souvent issus de partenariats entre certaines entreprises/instituts et l'école Polytechnique.

Keywords : Energy, Climate, Case-study.

2 Recherche Opérationnelle

Enseignants référents : Xavier Allamigeon Xavier.Allamigeon@inria.fr,
Stéphane Gaubert Stephane.Gaubert@inria.fr

Les projets de recherche opérationnelle peuvent être classés en deux grandes catégories : sujets portant sur les principales classes d’application de la recherche opérationnelle (énergie, télécom, transport, centre d’appels, tarification, . . .), d’une part, et d’autre part sujets d’initiation à la recherche, d’orientation fondamentale, portant sur des questions d’optimisation, de combinatoire, de complexité, et parfois aussi sur des thèmes d’interface (géométrie discrète ou tropicale, contrôle, théorie des jeux, modélisation). Bien entendu, cette répartition n’est pas étanche, certains sujets appliqués peuvent receler des aspects fondamentaux, et vice versa.

Des sujets sont proposés tous les ans par les enseignants de RO : XA, SG, Frédéric Meunier (ENPC) et Éric Gourdin (Orange Labs). Nous sollicitons aussi régulièrement un nombre limité de projets auprès d’autres chercheurs travaillant dans l’industrie (EDF, Air-France, Localsolver, . . .) Le travail de projet combine le plus souvent l’étude d’un article de recherche et une expérimentation numérique (test d’algorithmes). Les projets amont peuvent aussi comprendre des aspects théoriques originaux. On donne plus bas des exemples de projets des années précédentes – ce qui donne une idée des sujets susceptibles d’être proposées, sachant que les projets de RO sont orientés recherche : la plupart des projets sont originaux et sont renouvelés cette année.

Nous *recommandons fortement* aux élèves potentiellement intéressés de *contacter les coordinateurs* de la thématique (XA et SG), de préférence en amont du remplissage de votre fiche de choix, en précisant vos motivations scientifiques et vos souhaits de sous-thématiques. Nous recommandons en particulier de lister dans votre message des titres de projets des années précédentes qui vous auraient intéressés. Cela nous aide à apprécier vos intérêts afin de vous proposer des sujets répondant au mieux à vos attentes.

Il est aussi fortement recommandé qu’au moins un élève par binôme suive le cours MAP557 “Recherche Opérationnelle : Aspects mathématiques et applications” pour choisir ces projets.

Outils mathématiques : polyèdres, optimisation combinatoire, optimisation convexe, algorithmes de graphes, inégalités polynomiales (optimisation semi-algébrique), processus de décision Markoviens, jeux déterministes et stochastiques, géométrie discrète ou tropicale.

Classes d’applications : Énergie, télécom, transports, services d’urgence (centres d’appels, urgences hospitalières).

Les sous-thématiques pour la Recherche Opérationnelle sont :

- Applications de la recherche opérationnelle (énergie, transports, réseaux de communication, services d'urgence)
- Fondements mathématiques de la recherche opérationnelle.

2.1 Applications de la recherche opérationnelle

Des sujets relatifs à l'énergie et aux transports peuvent porter sur l'optimisation de production électrique, la gestion de flottes de véhicules électrique, des problèmes de tournées, ou de tarification, parfois avec des aspects touchant aux jeux de congestion. Des sujets issus du monde de l'aérien sont aussi susceptibles d'être proposés (planification des opérations ou tarification).

Une autre série de sujets porte sur l'application de l'optimisation combinatoire à des problèmes issus des réseaux télécom. Exemples de sujets donnés les années passées : répartition de serveurs virtuels sur des architectures physiques, optimisation des caches, ou contre mesures pour empêcher la propagation de virus dans un réseau.

Des sujets spécifiques portant sur le dimensionnement et l'évaluation de performance de centres d'appels d'urgences, en partenariat avec la direction de programme de la PFAU au sein de la préfecture de police (17-18-112), ou un service d'urgence de l'AP-HP, ont été proposés à plusieurs reprises. D'autres projets sur des centres d'appels commerciaux ont aussi été réalisés.

2.2 Fondements mathématiques de la recherche opérationnelle

Ces projets, destinés aux élèves désireux de s'initier à la recherche sous un angle théorique (par exemple, tentés par une thèse dans la discipline), portent sur l'analyse d'article récents, et parfois sur des prolongements. Des sujets à l'interface d'autres disciplines (géométrie discrète ou tropicale, contrôle optimal, dynamique des populations, algorithmique des jeux, dynamique épidémiologique) ont aussi été donnés. Donnons quelques thèmes explorés dans des projets récents : contre-exemples en programmation linéaire (montrant que les méthodes de points intérieurs peuvent effectuer un nombre d'opérations arithmétiques exponentiel en le nombre de variables et de contraintes), complexité des jeux de maximisation d'entropie (classe de jeux infinis avec paiement moyen), optimisation sous contraintes d'inégalités polynomiales, approximations polyédrales de dynamiques polynomiales (en cinétique chimique ou épidémiologie).

Liste indicative de sujets donnés les années précédentes

1. Dimensionnement couplé d'une flotte de véhicules électriques et des infrastructures de recharge
2. Planification de maintenance aérienne sous contrainte de ressource

3. “Tail assignment” (affectation des avions à des vols) avec contrainte de potentiel.
4. Approche robuste pour le dimensionnement capacitaire d’un réseau
5. Jeux de congestion appliqués au pilotage de la demande électrique
6. Planification d’investissement à long terme (par de production électrique).
7. Tarification de contrats électriques
8. Modèles pour le placement optimal dans les réseaux 5G
9. Optimisation de la réponse d’un centre d’appel d’urgence
10. Optimisation d’un service d’accueil des urgences
11. Dimensionnement d’un centre d’interventions de techniciens sur un réseau de distribution électrique
12. Dimensionnement et gestion d’une flotte de véhicules autonomes
13. Optimisation d’une organisation de ligne de production par “bucket brigades” (sujet d’ordonnancement à motivation appliquée et avec une forte composante fondamentale/combinatoire).
14. Algorithmes en transport optimal quantique
15. Barrières exotiques et complexité de la programmation linéaire
16. Complexité des jeux répétés avec paiement moyen : un approche basée sur la géométrie tropicale
17. Algorithmes pour les jeux d’entropie
18. Application de la géométrie tropicale en apprentissage et en analyse phylogénétique.
19. Géométrie discrète appliquée en économie (enchères, tarification).
20. Approximation polyédrale de dynamiques polynomiales en cinétique chimique ou épidémiologie.

3 Modèles aléatoires pour la physique, l'informatique et la biologie

Enseignant référent : Lucas Gerin lucas.gerin@polytechnique.edu

Pour cette thématique nous visons à décrire/simuler/prédire/comprendre des phénomènes aléatoires apparaissant en biologie ou en physique théorique. On peut aussi chercher à utiliser certains de ces phénomènes aléatoires en informatique pour concevoir des algorithmes efficaces, ou pour mieux comprendre certaines structures de données.

Un enjeu important pour beaucoup de modèles est de comprendre comment certains phénomènes peuvent émerger à différentes échelles : comment un modèle microscopique peut se décrire à une échelle macroscopique ?

Suivant vos goûts et les disponibilités nous vous aiguilleront vers un encadrant·e particulier (enseignant·e-chercheur·e au Département de Mathématiques Appliquées ou chercheur·se CNRS). Pour les thématiques 3.1 et 3.2 nous avons dans l'équipe PEIPS du laboratoire CMAP des spécialistes des domaines suivants :

- Probabilités et EDP pour l'évolution du vivant
- Processus stochastique (processus de branchement, processus en environnements aléatoires,...)
- Percolation et mécanique statistique
- Graphes et réseaux aléatoires
- Systèmes de particules en interaction, Systèmes complexes
- Combinatoire et structure discrètes
- Algorithmes probabilistes
- ...

Suivant le sujet et l'orientation que vous souhaitez donner au sujet les outils de mathématiques appliquées seront par exemple parmi :

- Chaînes de Markov, Martingales
- Marches aléatoires, processus stochastiques
- Équations différentielles, Systèmes dynamiques
- Graphes aléatoires, matrices aléatoires
- Combinatoire
- ...

Dans tous les cas le projet aura une forte composante Modélisation/Simulation. Il peut également y avoir une partie d'étude de l'état de l'art, ce projet peut être une excellente occasion de s'initier à la recherche en vue d'un M2/Thèse.

3.1 Modèles aléatoires pour la biologie

Nous développons dans l'équipe PEIPS du laboratoire CMAP des modèles aléatoires et déterministes pertinents pour capturer des phénomènes liés à la biodiversité, l'écologie et l'évolution. Plus spécifiquement, nous considérons des systèmes complexes construits essentiellement sur les comportements individuels (cellules, bactéries, espèces, populations, métapopulations) et qui prennent en compte le plus possible la biologie du problème.

La modélisation aléatoire permet de quantifier les fluctuations dont les sources peuvent être diverses : dérive génétique, aléa de mutations au moment de la reproduction, aléa des changements environnementaux, aléa dans le déplacement d'individus (impact de la fragmentation de l'habitat). La modélisation déterministe offre un point de vue plus macroscopique dans lequel les comportements individuels s'intègrent dans une évolution décrivant le comportement global du système.

Exemples de sujets proposés les années précédentes :

- Dilemme du prisonnier : la sélection produit-elle de la coopération ?
- Modèles de polymérisation en environnement fluctuant
- Modélisation de la propagation d'une épidémie de tumeurs contagieuses chez le diable de Tasmanie
- Descente de l'infini pour des modèles compétitifs
- ...

Pour les projets en P2, certains sujets peuvent être liés au cours MAP563 *Modèles aléatoires en écologie et évolution* si le binôme suit ce cours.

3.2 Modèles aléatoires pour la physique ou l'informatique

Une autre thématique développée dans l'équipe PEIPS concerne l'étude de grandes structures aléatoires discrètes et de leurs limites (volume infini ou limites d'échelles). On retrouve en effet au cœur de plusieurs problèmes naturels des objets discrets aléatoires fondamentaux comme des arbres, graphes, cartes planaires, permutations, systèmes de particules, etc. Il se trouve que ces objets trouvent à la fois des applications profondes en physique et en informatique.

Exemples de sujets proposés les années précédentes :

- Un modèle de milieu aléatoire en physique : la percolation de premier passage
- Inférences sur des arbres de Galton-Watson et fraude sur Wikipedia
- Simulations de processus déterminantaux
- Factorisations aléatoires de permutations
- Limites "champ-moyen" pour des graphes aléatoires *sparse*
- Propagation de rumeurs dans des graphes à attachement préférentiel
- ...

4 Modélisation et ingénierie financière

Enseignant référent : Stefano De Marco stefano.de-marco@polytechnique.edu

Il est obligatoire de suivre le cours MAP 552 Modèles stochastiques en finance pour choisir cet EA. Il est conseillé de suivre le cours MAP 556 Méthodes de Monte Carlo.

Outils : machine learning, optimisation et contrôle stochastique, calcul stochastique, processus aléatoires, statistiques des processus, théorie des jeux stochastiques, transport optimal.

Vous pouvez choisir parmi les sous-thématiques suivantes :

1. Évaluation et gestion des risques ;
2. Interactions stratégiques et risque systémique ;
3. Statistique haute fréquence et microstructure des marchés.

Suivant les possibilités d'encadrement et vos choix, nous vous aiguilleront vers un encadrant particulier (enseignant-chercheur au Département de Mathématiques Appliquées ou IP Paris).

Tous les projets auront une forte composante Modélisation/Simulation.

4.1 Évaluation et gestion des risques

La quantification et la gestion des risques est un problème primordial en finance de marché. Une question fondamentale est liée à la *couverture* des risques : lorsqu'une institution financière a pris des positions sur un marché qui vont engendrer des gains ou des pertes aléatoires (concrètement : on a construit un portefeuille d'actions, d'obligations ou de produits dérivés), comment peut-on réduire le risque de pertes futures ? Quantitativement, le problème consiste à contrôler l'espérance et la variance des pertes, et plus en général à contrôler leur distribution. C'est une question qui ouvre la porte à la modélisation aléatoire, et à la formulation de problèmes d'optimisation en environnement stochastique.

Les problèmes spécifiques que l'on peut aborder sont la couverture du risque sur les options (*option hedging*), dont la théorie est traitée dans le cours MAP 552, ou la calibration d'un modèle paramétrique à des données de marché. Mathématiquement, ces questions poussent à explorer les liens existant entre équations aux dérivées partielles et modèles stochastiques construits à partir du mouvement Brownien, et les méthodes de simulation Monte Carlo pour ces modèles. Les questions de complexité algorithmique (souvent liée au coût d'une simulation Monte-Carlo intensive) mènent parfois à faire recours à des algorithmes d'apprentissage – concrètement, on essaye d'approcher les fonctions de prix et de couverture générées par un modèle paramétrique avec des réseaux de neurones profonds, qui

sont ensuite déployés pour des calculs plus rapides lorsque le modèle est utilisé en pratique (c'est le sujet du *deep pricing*).

Des tendances de recherche plus récentes visent à s'affranchir de l'usage d'un modèle paramétrique spécifique : on peut considérer alors des problèmes de *couverture robuste* des risques sur les options exotiques, qui font intervenir naturellement la théorie du transport optimal, ou des méthodes basées sur l'apprentissage non-paramétrique d'une fonction de couverture optimale – c'est le problème du *deep hedging*.

4.2 Interactions stratégiques et risque systémique

L'étude des phénomènes d'interaction sur les marchés financiers est très importante. Les transactions sont établies par l'interaction d'acheteurs et de vendeurs qui ont des comportements stratégiques. La gestion des risques par les entreprises est fondée sur la diversification des risques par le jeu de détentions mutuelles. Ces interactions créent des interdépendances entre tous les acteurs du marché qui peuvent se révéler dangereuses du fait des phénomènes de contagion. En effet, la défaillance d'une petite entité peut avoir des conséquences catastrophiques sur une grande partie du système, conduisant même à des phénomènes de crises.

Afin de mieux cerner ces phénomènes, nous utilisons des modèles de théorie des jeux dans un cadre de dynamique stochastique.

4.3 Statistique haute fréquence et microstructure des marchés

Le but de la thématique statistique haute fréquence et microstructure des marchés est de proposer une introduction aux problématiques de trading intraday se posant de manière quotidienne à l'ensemble des institutions financières. A cette échelle, une compréhension fine du processus de formation des prix est indispensable afin de mettre en place des stratégies de trading optimales. Il est ainsi crucial de revoir la modélisation statistique des données de prix et de tenir compte des dynamiques de liquidité ainsi que du mécanisme de carnet d'ordres et du phénomène de *market impact*. Mathématiquement, il est nécessaire d'adapter les outils standards du calcul et du contrôle stochastique à ce nouveau contexte, pour les rendre en particulier compatibles avec les techniques spécifiques de statistique des processus et de machine learning employées en trading haute fréquence. Nous nous intéresserons ainsi à plusieurs questions de trading optimale importantes illustrant les challenges mathématiques, statistiques et numériques posés par la microstructure des marchés.

5 Simulation Aléatoire

Enseignant référent : Clément Rey clement.rey@polytechnique.edu

Dans cette thématique, on s'intéresse au calcul de grandeurs probabilistes par simulation de variables aléatoires. On porte une attention particulière à l'étude de méthodes proposant une approche numérique efficace pour cette simulation. On étudie théoriquement et numériquement la convergence de ces méthodes. On peut s'intéresser, entre autre, à la simulation d'Équations Différentielles Stochastiques (EDS), de Processus Markoviens ou non, à l'introduction d'incertitude dans un modèle ou encore à des méthodes de Meta-modélisation. Dans des cas pratiques, ces approches permettent, par exemple, de calculer des espérances, des quantiles, simuler des lois (ou des approximations) ou encore résoudre des problèmes d'optimisation stochastique.

Suivant les possibilités d'encadrement nous vous aiguillerons vers un encadrant·e particulier (enseignant·e-chercheur·se au Département de Mathématiques Appliquées, chercheur·se CNRS ou chercheur·se extérieur).

Suivant le sujet, les outils de mathématiques appliquées pourront être :

- Chaînes de Markov, Martingales
- Calcul stochastique
- Gestion du risque

Dans tous les cas, le projet aura une composante Modélisation/Simulation.

5.1 Processus aléatoires

Dans cette sous-thématique, l'attention se porte sur des méthodes de simulation associées à l'étude de processus aléatoires. On étudie la simulation de modèles qui entrent en jeu dans des domaines comme la finance ou la biologie. Ces modèles peuvent décrire l'évolution d'une grandeur comme le mouvement d'une particule, la valeur d'un actif sur un marché financier ou encore une population. On peut, par exemple, se focaliser sur le calcul d'une espérance associée à ce modèle, à la loi d'une grandeur comme le temps d'atteinte d'un seuil par un processus ou encore le calcul de quantiles. On met en application des techniques pouvant être héritées des chaînes de Markov, des martingales ou du calcul stochastique. On explore également des domaines comme les EDS non linéaires, la simulation de processus non-Markoviens ou l'étude d'événements rares.

Mots clés :

- Calcul Stochastique
- Équations Différentielles Stochastiques
- Martingales
- Processus Markoviens

—

Références

- N. El Karoui, E. Gobet : Les outils stochastiques des marchés financiers : Une visite guidée de Einstein à Black-Scholes
- D. Lambert, B. Lapeyre : Introduction au calcul stochastique appliqué à la finance
- D. Revuz, M. Yor Continuous Martingales and Brownian Motion

5.2 Quantification d'incertitude

La Quantification d'incertitude repose sur la gestion de l'aléa d'un modèle. Par exemple, un modèle (aléatoire) possède des paramètres qui sont des quantités déterministes. Ces paramètres ne sont, en réalité, pas toujours connus, car non observables ou non calculables exactement à partir de données d'observation. On peut alors introduire un aléa sur ces paramètres et obtenir un modèle étendu. On étudie alors, théoriquement et numériquement, l'impact de l'introduction de cet aléa dans la résolution de problèmes comme le calcul d'une espérance, de quantiles ou encore la résolution d'un problème d'optimisation. Une autre approche dans ce domaine repose sur la simulation même d'un modèle. Celle-ci étant parfois trop coûteuse, on peut introduire une approximation du modèle plus simple à simuler et dont on souhaite quantifier l'erreur associée lors de la résolution d'un problème comme ceux mentionnés. Ces méthodes, dites de Meta-modélisation, ainsi que celles évoquées en premier lieu, sont particulièrement utilisées dans la gestion du risque. On applique, entre autre, des techniques de probabilités et calcul stochastique combinées avec des techniques d'approximation de variables aléatoires (fonction d'une Gaussienne pour un exemple simple) par des projections, intrepolations, représentations en chaos, etc

Mots clés :

- Gestion du risque
- Modélisation et Meta-modélisation
- Représentation de variables aléatoires (*e.g.* Polynômes de chaos).
- Processus aléatoires
- ...

Références

- T. J. Sullivan : Introduction to Uncertainty Quantification
- J. Garnier : Cours de Gestion des incertitudes et analyse de risque (MAP 568)
- G. Szego : Orthogonal Polynomials

6 EDP, analyse numérique et applications en physique, mécanique et biologie

Enseignant référent : Teddy Pichard teddy.pichard@polytechnique.edu

Cette thématique concerne assez généralement les EDP et EDO issues de diverses modélisations en physique, mécanique ou en biologie. On s'intéressera en particulier aux méthodes numériques (déterministes) pour résoudre ces équations ou des problèmes qui leur sont liés.

Les sujets incluent généralement une ou plusieurs des composantes suivantes :

- Analyse des EDO/EDP/...
- Analyse numérique
- Calcul scientifique et calcul haute performance
- Optimisation (déterministe) avec ou sans contraintes
- ...

Dans tous les cas, le projet aura une forte composante modélisation et simulation déterministe. Selon la sous-thématique choisie, les cours MAP412, MAP431, MAP435 en 2A peuvent être recommandées.

Suivant les possibilités d'encadrement, nous vous aiguilleront vers un·e encadrant·e particulier·e (enseignant·e-chercheur·e au Département de Mathématiques Appliquées ou associé·e). Si vous êtes intéressé par un problème particulier relevant de cette thématique, ou souhaitez des précisions sur les éventuels sujets proposés, vous pouvez contacter par mail l'enseignant référent.

6.1 Sous-thématique : Modélisation et simulation

Domaines possibles d'applications :

- Équations de transport (transport de particules, radiation, ...)
- Mécanique des fluides (diphasique, sprays, ...)
- Modèles issues de la biologie (proie-prédateurs, diffusion croisée, ...)
- ...

6.2 Sous-thématique : Analyse numérique et calcul scientifique

Domaines possibles d'applications :

- Schéma numérique sous contraintes (préservant une asymptotique, la positivité, l'admissibilité, les équilibres...)
- Calcul haute performance (multi-résolution, algorithmes pararéels)
- ...

6.3 Sous-thématique : Optimisation déterministe

Domaines possibles d'applications :

- Optimisation de forme
- Optimisation sous contrainte (de modélisation)
- Transport optimal
- ...