Projets 3A

Projet 3ème année du PA de Mathématiques Appliquées

Catalogue des sujets*

26 mai 2016

*La dernière version de ce document se trouve sur la page MAP501 du catalogue Moodle : https://moodle.polytechnique.fr/enrol/index.php?id=3121

Table des matières

Introduction

Objectifs, rapport et soutenance, notation

▷ **Objectifs :**

Le but de cet enseignement 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, dans le but de proposer une réponse adaptée à une question de modélisation ou autour des techniques de calcul.

Ce projet personnel est **effectué en binôme** et constitue un véritable travail d'équipe.

Il n'est pas nécessairement associé à l'un des cours MAP de la période 1 ou période 2, mais les compétences de ces cours seront mises en jeu.

▷ **Rapport et soutenance :**

L'état d'avancement du travail sera évalue une première fois lors d'une **soutenance de mi-parcours** sur slides (20')

dans la semaine du 5 décembre 2016.

La remise d'un rapport avant cette soutenance de mi-parcours n'est pas obligatoire.

Le projet se concluera par la remise d'un **rapport**

dans la semaine du 13 mars 2017

à envoyer sous forme d'un fichier pdf à tous les membres du jury de soutenance (avec en cc l'enseignant responsable de la thématique du projet : voir plus loin pour la liste des thématiques), et à transmettre en deux copies papier au secrétariat du département MAP. Il fera l'objet d'une **soutenance** orale sur slides (40') à la fin de la période 2

dans la semaine du 20 mars 2017.

Les date exactes et les horaires des soutenances vous seront transmis par le département.

Dans le rapport comme lors de la soutenance, vous devez considérer que le jury ne connaît rien au problème, et donc le présenter en montrant son importance, expliquer votre approche ainsi que le cadre théorique, montrer évidemment vos résultats, et donner enfin vos conclusions sur le sujet en faisant un bilan de votre travail. Une bonne ligne de conduite est de faire comme s'il s'agissait d'une étude que l'on vous aurait commandée, à présenter aux commanditaires.

Le rapport, de préférence écrit en LaTeX, devra être rédigé soigneusement, et comprendre une bibliographie des ouvrages et articles étudiés. Vous êtes encouragés à chercher de la documentation sur votre sujet.

La soutenance orale finale dure 40 minutes par binôme : un exposé de 30 minutes (partagées équitablement par les élèves) suivi de 10 minutes de questions. Elle doit tenir dans le temps imparti.

▷ **Evaluation :**

Elle tiendra compte de la qualité du contenu et de la présentation 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 au deux membres du binôme.

Tous les sujets proposés demanderont d'effectuer des simulations et/ou du calcul numérique.

Une analyse critique des résultats ainsi obtenus devra être faite dans le rapport et la présentation orale. L'informatique devra être utilisée comme outil de compréhension et d'expérimentation, et non comme un fin en soi.

Thématiques

▷ **Liste des thématiques :**

La liste des sujets proposés est découpée dans les thèmatiques suivantes (en *italique* le référent pour chaque thématique) :

1. **Mécanique, physique, sciences de l'ingénieur** *(Aline Lefebvre-Lepot)*
2. **Data science** *(Stefane Gaïffas)*
   (répartie en sous-thématiques : son et image, réseaux et graphes, statistique et apprentissage)
3. **Optimisation et recherche opérationnelle** *(Xavier Allamigeon, Stéphane Gaubert)*
4. **Sciences de la vie** *(Lucas Gerin)*
5. **Mathématiques financières** *(Stefano De Marco, Nizar Touzi)*
   (répartie en plusieurs sous-thématiques : voir catalogue)

▷ **Enseignants référents pour les différentes thématiques :**

Pour chacune des thématiques, un enseignant référent est disponible pour répondre à vos questions. Ses coordonnées sont indiquées dans la suite de ce

catalogue, au début de la section pour chaque thématique. **N'hésitez pas à prendre contact avec lui pour tout souci pendant la durée du projet** (problème d'organisation, problème scientifique...).

Choix et attribution des sujets

▷ **Procédure de choix :**

Via le Moodle MAP501 : soumettre dans l'Assignement "Choix des projets" ouvert sur le Moodle

- 4 choix de sujets classés par ordre de préférence
- en indiquant le numéro de sujet (1.2, 3.4, 2.1.8...)
- dans au moins 2 thématiques différentes
- Pour les thématique Mathématiques financières et Data science : si 3 choix sont indiqués, ils doivent etre répartis sur au moins deux sous-thématiques.

Choix à exprimer

entre le 26/05/2016 à 10h et le 9/06/2016 à 10h

Une seule soumission par binome suffit. Indiquez clairement le nom de votre binome avec le choix de sujets.

▷ Articulation avec les EA MAP571 : il ne sera pas possible d'effectuer son projet 3A et son EA MAP571 dans la même thématique.

▷ **Attribution des sujets :**

Elle sera faite au plus vite, en essayant de respecter au mieux les choix envoyés à l'heure et respectant les règles de choix ci-dessus. Ceux-ci seront tous traités de façon égalitaire.

Les autres seront traités ensuite, et par ordre d'arrivée.

Déroulement du projet, encadrement

▷ **Démarrage du projet :**

Quand votre sujet vous aura été attribué, vous devrez prendre contact avec l'enseignant qui en est responsable avant le 13 septembre 2016 afin de fixer la date d'un premier rendez-vous.

Cet enseignant vous précisera les modalités de travail particulières pour votre sujet, puis vous encadrera.

▷ **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 à lui de déterminer les modalités de vos rencontres
(et non l'inverse).

Ces rencontres n'auront pas nécessairement lieu le mardi, mais les créneaux libres ces jours-là, aussi que les autres créneaux à disposition dans l'emploi du temps pour travailler votre projet 3A, doivent être utilisés à bon escient.

Bon travail à tous.

Liste des sujets

1 Mécanique, physique, sciences de l'ingénieur

Enseignant référent : Aline Lefebvre-Lepot aline.lefebvre@polytechnique.edu

1.1 Controlling Quantum Mechanical Systems

Ugo Boscain ugo.boscain@polytechnique.edu

A control system is a dynamical system on which one can act from outside via certain parameters that one can vary in time. Typical examples are for instance robots and satellites (indeed almost every dynamical system that one meets in engineering can be modelled as a control system). In mathematical terms one is faced with a problem of the following type :

$$\dot{x} = f(x(t), u(t)), \qquad x(0) = x_0, \quad x \in \mathbf{R^n}, \quad u(.) : \mathbf{R} \to U \subset \mathbf{R^m}. \qquad (1)$$

Typical questions that one meet in control theory are the following :

- for every x_1 is it possible to find T and $u(.)$ such that the solution of (1) satisfy $x(T) = x_1$? If not, which are the points that can be reached ?
- is it possible to solve the previous question for an arbitrarily small T?
- find the control such that the corresponding trajectory steer x_0 in x_1 minimizing a cost. For instance a energy like cost as $\int_0^T |u(t)|^2\, dt$.

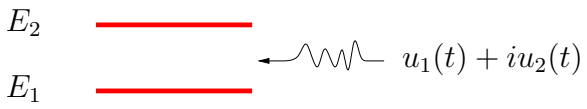
Today one of the most striking application of control theory is quantum mechanics. Typical examples are

- computation of magnetic fields in nuclear magnetic resonance for medical imaging,
- realization of quantum gates in quantum computers,
- induction of chemical reaction via external fields (photochemistry).

The purpose of this projet is to study some general fact of control of quantum mechanical systems. Beside a theoretical part, the students will learn how to simulate and control a spin 1/2 particle that is the most simple, but most important controlled quantum mechanical system in applications.

$$i\dot{\psi}(t) = \begin{pmatrix} E_1 & u_1(t) + iu_2(t) \\ u_1(t) - iu_2(t) & E_2 \end{pmatrix} \psi(t). \qquad (1)$$

Here $\psi(\cdot) = (\psi_1(\cdot), \psi_2(\cdot)) : [0, T] \to \mathbb{C}^2$ describes the state of the system evolving on the unit sphere of $\mathbb{C}^2$. The quantities $E_1, E_2 \in \mathbf{R}$ represent the energy levels of the system. The controls $u_1(\cdot)$ and $u_2(\cdot)$, that are functions from $[0, T] \to \mathbf{R}$, describe the action of the external fields.



To choose this subject, it should be better but not necessary to follow MAP561 (Automatic) during the second period.

Références

[1] U. Boscain et Y. Chitour. Notes de cours Edition 2015/2016 MAP 561. Introduction Introduction to Automatic Control.
http ://www.cmapx.polytechnique.fr/∼boscain/AUTOMATICS/
[2] D. D'Alessandro, Introduction to quantum control and dynamics. Applied Mathematics and Nonlinear Science Series. Boca Raton, FL : Chapman, Hall/CRC., 2008
[3] U. Boscain and P. Mason, Time minimal trajectories for a spin 1/2 particle in a magnetic field, J. Math. Phys. 47, 062101 563 (2006).

1.2 Time Optimal Control for a UAV Drone

Ugo Boscain ugo.boscain@polytechnique.edu

We consider the problem of controlling an unmanned aerial vehicle (UAV) flying at a constant altitude (HALE type) to provide a target supervision. We make the following assumptions on the UAV :

- the velocity of the UAV is assumed to be constant ;
- the UAV is assumed to be kinematically restricted by its minimum turning radius $r > 0$, or equivalently, its yaw angle is assumed to be constrained by an upper positive bound.

The UAV is modeled as a Dubins vehicle (i.e. a planar vehicle with constrained turning radius and constant forward velocity, see [1]).

$$\begin{cases} \dot{x} = \cos\theta \\ \dot{y} = \sin\theta \\ \dot{\theta} = u(t). \end{cases} \tag{2}$$

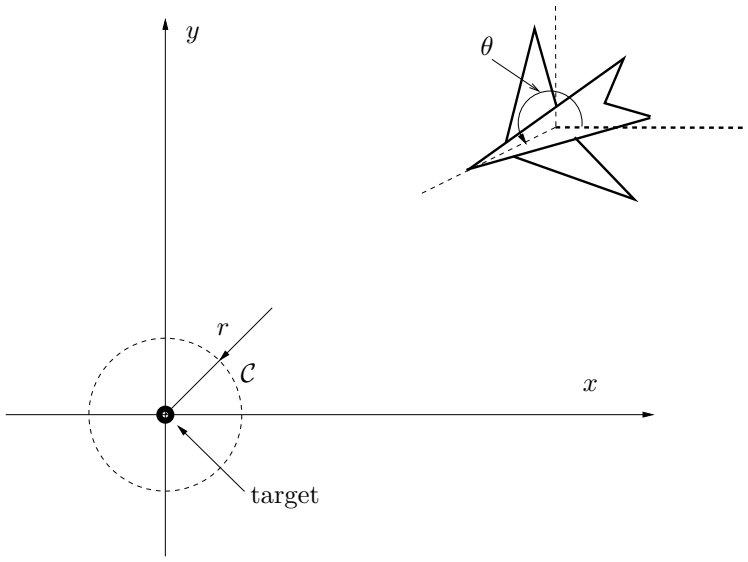
with $(x, y, \theta) \in \mathbf{R}^2 \times \mathbb{S}^1$ being the state (where $(x, y) \in \mathbf{R}^2$ is the UAV's coordinate in the constant altitude plane, and θ the yaw angle), and $u : [0, T] \to [-u_{\max}, u_{\max}]$ being the control variable.

Note that the yaw angle θ is the angle made by the aircraft direction with respect to the x-axis. These equations express that the drone evolves on a perfect plane (perfect constant altitude), at perfect constant speed 1, moves in the direction of its velocity vector, and is able to turn right and left with a minimal turning radius $r = 1/u_{\max}$.

The purpose here is to find the time-optimal trajectory tracking the UAV from its initial position (x_0, y_0, θ_0) to a circle of minimal radius centred on the target

(that is assumed to be placed at the origin) :

$$\mathcal{C} = \{(x, y, \theta) \mid x = r\sin\theta,\ y = -r\cos\theta\}\,.$$



To choose this subject, it should be better but not necessary to follow MAP561 (Automatic) during the second period.

Références

[1] [1] A. A. Agrachev and Y. L. Sachkov. Control theory from the geometric viewpoint, volume 87 of Encyclopaedia of Mathematical Sciences. Springer-Verlag, Berlin, 2004. Control Theory and Optimization, II.

1.3 Control of a UAV Formation

A. El Hadri (elhadri@lisv.uvsq.fr), A. Benallegue (benalleg@lisv.uvsq.fr)

Nowadays, the use of UAV is steadily growing and a lot of applications can be envisaged. Some applications such as surveillance and searching objects or large payloads transportations require the cooperation of multiple UAVs. In this case, moving into formation is considered. The planning and design of trajectories for multiple UAVs depends on strategy of formation control according to different kinds of scenarios. Flying in formation means that the distances between individual pairs of UAVs stay fixed and the formation of UAVs can be considered moving as a rigid entity. A formation of multiple UAVs can be modeled as a multi-agent system.

The problem to be addressed is how to characterize the choice of agent pairs to preserve the shape property of the formation and which strategy of control will be used to stably maintain or restore the shape of a formation. The use of graph theory, nonlinear systems theory and linear algebra is relevant for such problem.

We consider a practical problem of flying a group of three UAVs in an equilateral triangle that are involved in a cooperative task to move a payload from point A to point B (Figure 1). The control of a formation requires consideration of several aspects such as moving the centre of mass of the formation by adopting a certain orientation ; maintain the relative positions of the agents during movement, so that the form remains preserved, avoid obstacles, etc. Indeed, the highest level problem is how to define practical architecture for the formation while it moves as a cohesive whole.

FIGURE 1 – Formation of three UAVs for payload transportation task

For the control of a triangular formation, we assume that the formation exists in the plane and agents are point agents. Denote the three UAVs (agents) by 1, 2 and 3, and suppose their positions at any instant of time are denoted by x_i ; $i = 1; 2; 3$. Suppose the nominal distances from 1 to 2, 2 to 3 and 3 to 1 are d_{12} ; d_{23} ; d_{31} satisfying the triangle inequality. Let z_1 ; z_2 ; z_3 denote the relative positions of 1 with respect to 2, etc, i.e. $z_1 = x_1 - x_2$. The formation stabilization task is to ensure that $lim_{t\to\infty}$ $(\|zi(t)\| - di) = 0$; for $i = 1; 2; 3$. The question is, such behavior can it be reasonably guaranteed for all initial conditions and which control can ensure that.

The students can investigate different approaches for autonomous formations and shape maintenance. They can examine approach with symmetric control structure represented with undirected graphs, or approach with asymmetric control structure represented with directed graphs. A decentralized control laws can be used to ensure that the shape of a formation is preserved. The architecture for the formation can be done with leader and follower or by three-coleaders.

For payload transportation task, we consider a path to be followed by the center of mass (CM) of the formation (see 2) :

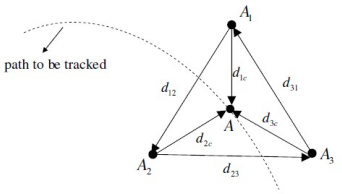


FIGURE 2 – Formation structure and path tracking for payload transportation task

To tracking perfectly the path, we can consider a virtual point agent A of the

formation with kinematic given by

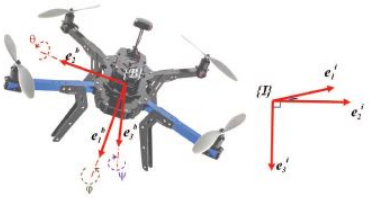
$$\begin{cases} \dot{p}_A(t) & = v_A(t), \\ \dot{R}_A(t) & = R_A(t)S(\omega_A(t)), \end{cases}$$

where $p_A(t)$ and $v_A(t)$ represent the position and velocity of the virtual point agent A and $R_A(t)$ represents its orientation.

Then, matching the virtual agent kinematics with the desired path allows to design a control scheme for the 3-UAV to maintain the equilateral rigid formation while the CM of this formation is tracking the virtual agent A.

Each agent A_i for $i = \{1, 2, 3\}$, can be modeled using kinematics and dynamics of a rigid body as

$$\begin{cases} \dot{p}_i(t) & = v_i(t), \\ m_i\dot{v}_i(t) & = -T_i(t)R_i(t)e_3 + m_i g e_3 \\ \dot{R}_i(t) & = R_i(t)S(\omega_i(t)), \\ J_i\dot{\omega}_i(t) & = -S(\omega_i(t))J_i\omega_i(t) + \tau_i(t) \end{cases}$$



UAV- Rigid Body

where $p_i(t)$ and $v_i(t)$ are the position and velocity of the UAV (agent A_i). $R_i(t)$ is its attiude and $\omega_i(t)$ is the angular velocity in body frame. m_i is the masse and $J_i$is the inertial moment of the UAV (i). $\tau_i(t)$ is the control torque applied to the UAV (i) and T_i represents the magnitude of the total thrust of each UAV.

All suitable approaches should be validated first by simulation. Then an experimental device consisting of theree UAV Solo of 3Dr Robotics [1] is used for real tests. Each Solo UAV is equipped with Pixhawk autopilot and the needed sensors and network communication.

References

1- B. Anderson, C. Yu, S. Dasgupta, and A. Morse, "Control of a three coleaders formation in the plane", Systems & Control Letters, vol. 56, pp. 573-578, 2007.

2- R. Olfati-Saber and R. M. Murray, "Distributed cooperative control of multiple vehicle formations using structural potential functions", in Proc. of the 15th IFAC World Congress, (Barcelona, Spain), pp. 1-7, 2002.

3- R. Olfati-Saber and R. M. Murray, "Graph rigidity and distributed formation stabilization of multi- vehicle systems", in Proc of

1. https ://3dr.com/solo-drone/

the 41st IEEE Conf. on Decision and Control, (Las Vegas, NV), pp. 2965-2971, 2002.

4- B. Anderson, B. Fidan, C. Yu, D. Walle, UAV formation control : theory and application, in : V. Blondel, S. Boyd, H. Kimura (Eds.), Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, Vol. 371, Springer, Berlin/Heidelberg, 2008, pp. 15–33.

5- Zhicheng HOU, “Modeling and formation controller design for multi-quadrotor systems with leader-follower conguration”, Phd thesis of Université de Technologie de Compiègne, Laboratoire Heudiasyc UMR CNRS 7253, 10 Février 2016

1.4 Simulation numérique pour la condensation de Bose Einstein

Anne de Bouard debouard@cmap.polytechnique.fr
Romain Poncet romain.poncet@cmap.polytechnique.fr

En reprenant les travaux de Satyendra Nath Bose, Albert Einstein prédit en 1925 qu'un gaz bosonique parfait subit une transition de phase pour des températures de l'ordre de quelques nanokelvins, si sa densité devient suffisamment grande. Ce changement de phase mène à la formation d'un agrégat macroscopique de bosons, régi par une dynamique quantique : c'est le condensat de Bose Einstein.

Ce n'est que depuis une vingtaine d'années que les physiciens sont capables de réaliser ces condensats, par le biais d'expériences extrêmement complexes. Ce projet propose d'étudier quelques méthodes numériques qui permettent de simuler la dynamique de tels objets, en partant de l'équation de Gross-Pitaevkii décrivant la fonction d'onde du condensat, dans le but de retrouver numériquement des résultats expérimentaux récents établis dans la littérature physique. En fonction de leurs intérêts, les étudiants pourront par exemple mettre en évidence la formation spontanée de vortex dans des condensats en rotation, ou calculer des états stationnaires de ces systèmes, comme en Figure 1.4.

Ce projet permettra aux élèves de découvrir les modélisations physiques de ces états de la matière, ainsi que les méthodes numériques utilisées pour la résolution de ces modèles.

Pour choisir ce sujet, il serait préférable d'avoir suivi MAP411.

Références

[1] Antoine, Xavier, Weizhu Bao, and Christophe Besse. "Computational methods for the dynamics of the nonlinear Schrödinger/Gross-Pitaevskii equations." Computer Physics Communications 184.12 (2013) : 2621-2633.

[2] Tsubota, Makoto, Kenichi Kasamatsu, and Masahito Ueda. "Vortex nucleation and array formation in a rotating Bose-Einstein condensate." Journal of low temperature physics 126.1-2 (2002) : 461-466.

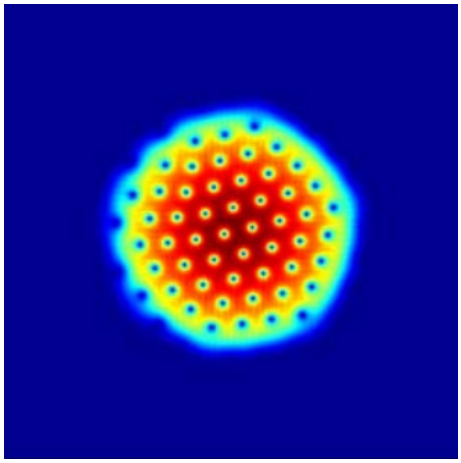


FIGURE 3 – Calcul d'un état stationnaire d'un condensat.

1.5 Simulation numérique en mécanique quantique

Eric Cancès cances@cermics.enpc.fr

La simulation numérique est aujourd'hui un outil de recherche très utilisé en mécanique quantique, qui vient en complément des approches purement théoriques et expérimentales. L'objet de cet enseignement d'approfondissement est d'étudier d'un point de vue mathématique et numérique quelques modèles couramment utilisés en physique et chimie quantique, ainsi qu'en science des matériaux et en biologie moléculaire. On pourra par exemple s'intéresser, en fonction des connaissances et des goûts des étudiants :

1. aux modèlles de Hartree-Fock et de Kohn-Sham (simulation moléculaire *ab initio*) ;
2. à l'équation de Schrödinger périodique (physique du solide, science des matériaux) ;
3. au modèle de Gross-Pitaevskii (condensats de Bose-Einstein) ;
4. au contrôle optimal de l'équation de Schrödinger dépendant du temps (contrôle de systèmes quantiques).

Références

[1] X. Antoine and R. Duboscq, *Modeling and Computation of Bose-Einstein Condensates : Stationary States, Nucleation, Dynamics, Stochasticity*, Lecture Notes in Mathematics/Physics, Springer 2015.
http ://becasim.math.cnrs.fr/publications2/files/AntoineDuboscqChapter.pdf

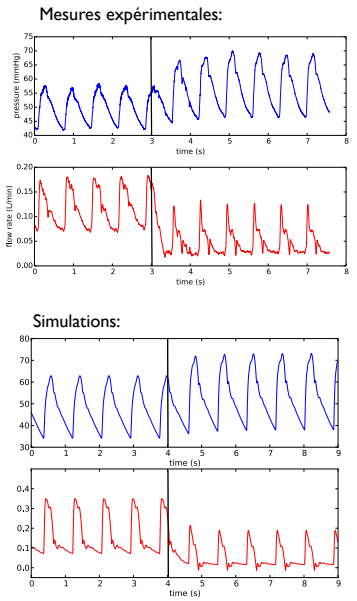
[2] E. Cancès, *Mathematical models and numerical methods for electronic structure calculation*, ICM Proceedings 2014.
http ://cermics.enpc.fr/∼ cances/ICM-Proceedings-Cances.pdf

[3] S.J. Glaser et al., *Training Schrödinger's cat : quantum optimal control*, Eur. Phys. J. D 69 (2015) 279.
http ://arxiv.org/abs/1508.00442

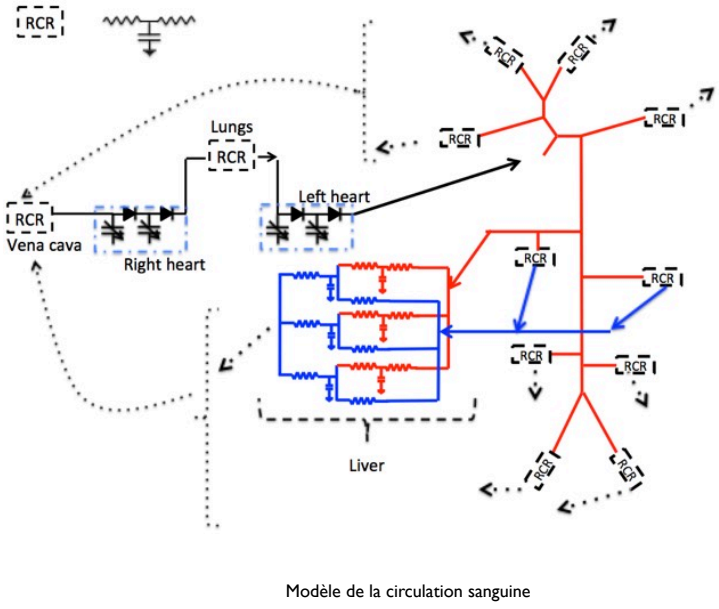
1.6 Régulation dans le système cardiovasculaire

Jean-Frédéric Gerbeau jean-frederic.gerbeau@inria.fr

On considère dans ce projet un modèle global de la circulation sanguine, incluant les circulations systémique et pulmonaire, ainsi que les quatre cavités cardiaques. Mathématiquement, le modèle est constitué d'équations aux dérivées partielles hyperboliques (systèmes de lois de conservation de type Euler) décrivant le débit et la pression le long de portions cylindriques de vaisseaux, et d'équations différentielles ordinaires représentant l'évolution en temps de la pression et du débit sanguin dans des compartiments vasculaires. Ce modèle permet de reproduire des observations expérimentales non triviales comme des changements de forme d'onde de pression suite à des opérations chirurgicales. Il ne prend cependant pas en compte les mécanismes de régulation présents dans la réalité.



Pression et débit dans l'artère hépatique, avant et après chirurgie



Modèle de la circulation sanguine

Ce projet a pour but d'enrichir le modèle de certains de ces mécanismes de régulation afin d'être capable de simuler l'impact cardiovasculaire de diverses situations

(changement de position, exercice, choc circulatoire, *etc.*). Le travail comportera une partie de modélisation, c'est-à-dire une traduction en équations de connaissances issues de la littérature biomédicale, et une partie de développements numériques dans un code C++.

Des connaissances de base des méthodes numériques pour les équations aux dérivées partielles suffisent pour aborder le projet, si elles sont accompagnées d'un fort intérêt pour le calcul scientifique et la modélisation biomédicale.

1.7 Réfraction négative et lentilles parfaites

Houssem Haddar Houssem.Haddar@polytechnique.edu

Pour des lentilles conventionnelles, la lumière est focalisée par la courbure des lentilles et la netteté des images est limitée par la longueur d'onde (de la lumière). Il s'agit ici de contraintes qui s'expliquent par la lois de Descartes en optique géométrique et la tâche focale de Rayleigh en diffraction. Ces théories classiques reposent sur le principe que l'indice de réfraction d'un milieu est toujours positif du fait que les constantes diélectriques sont positives. Or ce principe a été mis en défaut par de nombreuses expériementations (récentes) qui ont révélées que ces constantes dépendent des fréquences et que pour certaines fréquences elles peuvent avoir des valeurs négatives (notamment pour certains métaux). J.B. Pendry et ses collabrateurs ont aussi montré qu'un agencement périodique de résonateurs électriques et magnétiques (bien dimensionnés) pourrait engendrer un indice équivalent (macroscopique) négatif : on parle de métamatériaux. Ces démontrations expérimentales ont remis à l'ordre du jour les prédictions V. Veselgo 50 ans plus tôt sur certaines applications inattendues et révolutionnaires de ce type de matériaux, notamment en optique et la possibilité de fabriquer des lentilles non courbes lorsque l'indice vaut -1. J.B. Pendry montre aussi que dans ce cas les lentilles sont parfaites, dans le sens où elles ont une focalisation indépendante de la longueur d'onde [1]. Cette prédiction théorique d'une précision infinie ouvre bien évidemment la voie vers d'innombrables applications en imagerie/microscopie... Mais encore faut-il être capable de concevoir des métamatériaux qui ont un indice égal à -1 à toute fréquence [2]. Les matheux se sont également attelés à l'étude du problème de diffraction par des matériaux à indice négatif dans des configurations géométriques complexes afin de mieux comprendre la physique de ce type matériaux. L'indice négatif change aussi la «nature mathématique»du problème de diffraction et rend son étude théorique et numérique plus difficile [3].

Ce projet a pour but de s'initier à la physique des métamatériaux et de leurs applications ainsi qu'à l'étude théorique et numérique de l'expérience de diffraction négative dans le cas simplifié de couches diélectriques/métaux empilées. On essayera en particulier de mettre en évidence numériquement les phénomènes de réfraction négative et de super-résolution (dépassement de la limite de diffraction Rayleigh).

Prérequis souhaité : MAP431

Références
[1] J.B. Pendry, Negative Refraction Makes Perfect Lens, Physical Review Letters, v85, n8, 2000
http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.85.3966
[2] J.B. Pendry et D.R. Smith, The Quest For Superlens, Scientific American (67)
[3] L. Chesnel, Étude de quelques problèmes de transmission avec changement de signe. Application aux métamatériaux. Thèse,
https://tel.archives-ouvertes.fr/pastel-00763206/

1.8 Que peut-on extraire à partir d'une image réciproque ?

Houssem Haddar Houssem.Haddar@polytechnique.edu

En mettant un cheveu en face d'un pointeur laser et en observant la lumière diffractée sur un mur orthogonal à la direction du rayon, on observe des franges à espacement régulier. L'espacement des franges est directement lié à l'épaisseur du cheveu (on pourra commencer par établir théoriquement la formule liant l'épaisseur à la distance cheveu-mur et à la longueur d'onde du laser puis la valider expérimentalement !). On appelle image réciproque du cheveu le motif créé par le faisceau laser sur le plan orthogonal. Exploiter ces images réciproques pour obtenir une information quantitative sur la forme et la taille réelle d'un objet est par exemple la base de la cristallographie. En effet, le phénomène décrit pour les cheveux est reproductible à toute échelle de mesure si on effectue la même mise à l'échelle pour la longueur d'onde utilisée. C'est ainsi qu'en utilisant des rayons X, dont la longueur d'onde est de l'ordre du Angström, il est possible de caractériser la distribution de tailles de nanoparticules dans un échantillon (technique SAXS [1] - voir aussi l'instrument de mesure http://www.xenocs.com/). Le problème est cependant non linéaire et est fortement mal posé car trop sensible aux erreurs de mesures. Ce projet a pour but d'étudier la version linéarisable du problème (lorsque les interactions entre les particules est négligeable) et de comparer/améliorer différentes stratégies de régularisation du caractère mal posé du problème [2]. On s'intéressera en premier lieu au cas de mesures isotropes (correspondant à une distribution et des orientations aléatoires des particules). Le traitement d'anisotropie est une des ouvertures de ce projet : quel type d'anisotropie permet de traiter ce type de technique et quelles applications en imagerie de nanostructures peut-on envisager ?

Prérequis souhaité : MAP411

Références
[1] Elements of Modern X-ray Physics, Jens Als-Nielsen, Des McMorrow, Wiley, 2011
[2] A robust inversion method according to a new notion of regularization for

Poisson data with an application to nanoparticle volume determination Federico Benvenuto, Houssem Haddar, Lantz Blandine Siam J. Appl Math, 2015.

1.9 Adaptation de maillage et estimation d'erreur pour multiples fonctionnelles de but

Thomas Wick thomas.wick@ricam.oeaw.ac.at

Dans de nombreuses applications en physique et en ingénierie, l'objectif n'est pas seulement une solution approchée des équations aux dérivées partielles (EDP), mais le calcul d'un objectif fonctionnel spécifique. Par exemple, dans la dynamique des fluides, nous sommes intéressés par le calcul précis de forces et de moments. Les applications comprennent l'aéronautique, aérodynamique ou la météorologie.

Une méthode de choix est l'estimation d'erreurs spécifique à un but, utilisant des mesures locales d'un problème adjoint. Employant la méthode des éléments finis pour calculer la solution approchée d'une équation aux dérivées partielles, nous travaillons avec un maillage qui représente le domaine. Pour le calcul précis d'un objectif fonctionnel, le maillage est raffiné avec les informations obtenues par l'estimation d'erreur. Pour identifier les éléments qui doivent être raffinés, l'estimateur d'erreur doit localiser les zones d'erreur maximale. Une méthode récente repose sur une décomposition de l'unité qui est bien adaptée pour le traitement des équations aux dérivées partielles couplées et des problèmes multiphysiques comme par exemple les équations de Navier-Stokes combinée avec l'élasticité ou la mécanique de la rupture. En effet les derniers exemples pourraient demander pour le calcul des plusieurs fonctionnelles de but, par exemple une force et un déplacement simultanément.

Dans ce projet, nous récapitulons d'abord l'estimation d'erreurs pour un but d'une seule fonctionnelle linéaire et une équation aux dérivées partielles linéaire (par exemple le problème de Poisson). Par la suite, notre objectif est le développement de la méthode pour plusieurs fonctionnelles de but. Enfin nous considérons des équations non-linéaires (par exemple p-Laplace) et des problèmes vectoriels comme Navier-Stokes.

Ce projet se compose de développements algorithmiques qui seront implémentés dans la bibliothèque deal.II (C++).

Prérequis souhaité : MAP411 et/ou MAP431

Références

[1] R. Becker and R. Rannacher ; A feed-back approach to error control in finite element methods : basic analysis and examples, East-West J. Numer. Math., Vol. 4, 1996, pp. 237-264

[2] T. Richter, T. Wick ; Variational Localizations of the Dual-Weighted Residual Estimator, Journal of Computational and Applied Mathematics, Vol. 279 (2015), pp. 192-208

[3] R. Hartmann and P. Houston ; Goal-Oriented A Posteriori Error Estimation

for Multiple Target Functionals, In Hyperbolic Problems : Theory, Numerics, Applications, 2003, pp. 579-588

1.10 La méthode de Newton pour problèmes avec Hessienne indéfinie

Thomas Wick thomas.wick@ricam.oeaw.ac.at

La méthode de Newton est une approche performante pour trouver les racines des problèmes non linéaires. Dans ce projet, nous appliquons celle-ci pour un système d'équations aux dérivées partielles (EDPs) couplées et en particulier une formulation de la rupture variationnelle. Cette idée fut développée par Francfort et Marigo en 1998 et repose sur les lois classiques de Griffith. Nous en trouvons des applications dans des domaines aussi divers que la mécanique de la rupture dans les fissures solides, la fatigue et les défauts microscopiques des matéraux, les fissures dans le milieu poreux et multiphysique, l'énergie géothermique ou même des problèmes biomédicaux comme la dissection de l'aorte.

En travaillant avec une forme variationnelle, nous cherchons des solutions pour une variable de fissure (aussi connue comme champ de phase) et les déplacements. Une discrétisation simultanée des deux variables se traduit par un problème de minimisation numérique avec une Hessienne indéfinie. Bien que cette forme soit difficile à résoudre, nous pouvons nous appuyer sur plusieurs outils de l'optimisation numérique pour gérer ces déficiences comme par exemple une modification de la méthode de Newton avec un algorithme de recherche en ligne, ce qui permets une courbure négative.

Le but de ce projet est l'application de méthode Newton adaptée à la rupture variationnelle en développant un cadre numérique robuste. Comme première étape nous considérons un problème simplifié avec une Hessienne définie positive. Ensuite, le problème original sera traité. Ici le travail principal sera de déterminer les directions de la courbure négative qui comprend une analyse des valeurs propres. La bibliothèque fondamentale sera deal.II (C++).

Prérequis souhaité : MAP411 et/ou MAP431

Références

[1] G.A. Francfort and J.-J. Marigo ; Revisiting brittle fracture as an energy minimization problem, J. Mech. Phys. Solids, Vol. 46, 1998, pp. 1319-1342
[2] J. Nocedal and S. J. Wright ; Numerical optimization, Springer Ser. Oper. Res. Financial Engrg., 2006
[3] J.J. More and D.C. Sorensen ; On the use of directions of negative curvature in a modified Newton method, Mathematical Programming 16, 1979, pp. 1-20

1.11 Modélisation et identification de deux populations neuronales impliquées dans la maladie de Parkinson

Antoine Chaillet (CentraleSupélec – Univ. Paris Sud) antoine.chaillet@centralesupelec.fr

Contexte : La réponse typique d'un neurone à une stimulation électrique constante est caractérisée par des décharges de son potentiel de membrane. Ces réponses sont à la base de la communication neuronale et sont donc impliquées dans la plupart des mécanismes cérébraux. Ils peuvent également être à la base d'activités pathologiques : c'est le cas de la maladie de Parkinson, dont les symptômes moteurs sont fortement corrélés à une suractivité neuronale dans la bande de fréquences beta (13-30Hz). De récents travaux de la communauté des neurosciences proposent des modèles spatio-temporels permettant d'évaluer le taux de décharges d'une zone cérébrale donnée et de caractériser leur rôle dans la maladie de Parkinson. Ces modèles sont notamment utilisés dans le projet ANR SYNCHNEURO, dans le cadre duquel se déroulera ce projet, en collaboration avec des neurophysiologistes et neurochirurgiens de l'Hôpital H. Mondor de Créteil.

Objectifs : Ce projet vise à modéliser et identifier les paramètres de deux structures cérébrales impliquées dans la maladie de Parkinson (le noyau sous-thalamique et le globus pallidus externe). Cette identification se fera à partir de données expérimentales collectées in vivo en conditions saines et en conditions pathologiques. Ces données sont obtenues sous optogénétique, qui permet une activation neuronale commandée par des impulsions lumineuses. Il s'agira, pour chaque population neuronale, de déterminer les paramètres (fonction d'activation, constante de temps, retards, noyaux synaptiques,...) permettant de reproduire au mieux les comportements observés. A terme, ce modèle sera utilisé pour le développement de stratégies de stimulation lumineuse en boucle fermée.

Principaux points du projet :

- Bibliographie : familiarisation avec le modèle (neural fields), connaissance de base des mécanismes impliqués dans la maladie de Parkinson, techniques d'identification
- Implémentation numérique du modèle
- Identification paramétrique
- Comparaison des paramètres en conditions saines et pathologiques.

Domaines concernés : automatique non-linéaire, neurosciences, systèmes dynamiques.
Lieu : Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec
Perspectives : ce projet pourrait conduire à une thèse de doctorat

2 Data science : son & image, réseaux, apprentissage

Enseignant référent : Stéphane Gaïffas stephane.gaiffas@cmap.polytechnique.fr

Les projets liés à la science des données sont répartis dans les trois grands thèmes suivants :

- Signal, image et séries chronologiques
- Réseaux et graphes
- Statistique et apprentissage

Nous indiquons les cours associés à chaque thème, pour pour vous permettre de faire des choix cohérents avec les cours suivis.

Equipe enseignante pour les projets science des données

- Stephanie Allassoniere allassonniere@cmap.polytechnique.fr
- Emmanuel Bacry emmanuel.bacry@polytechnique.fr
- Olivier Cappe cappe@telecom-paristech.fr
- Stéphane Gaïffas stephane.gaiffas@polytechnique.edu (coordinateur des projets datascience)
- Julie Josse josse@agrocampus-ouest.fr
- Marc Lavielle marc.lavielle@inria.fr
- Marc Lelarge marc.lelarge@ens.fr
- Erwan Le Pennec erwan.le-pennec@polytechnique.edu
- Eric Matzner-Lober eml@uhb.fr
- Francois Roueff roueff@telecom-paristech.fr
- Erwan Scornet erwan.scornet@upmc.fr

2.1 Signal, image et séries chronologiques

Il est recommandé de suivre l'un des cours suivants pour travailler sur les projets proposés dans cette section :

- MAP555 - Signal processing
- MAP565 - Time series analysis

2.1.1 Séparation de sources : l'analyse en composantes indépendantes

Stéphanie Allassonnière stephanie.allassonniere@polytechnique.edu

L'analyse en composantes indépendantes (ACI) aussi appelée séparation de sources est une méthode d'analyse statistique de données. Elle permet de représenter ces données vectorielles sous la forme de combinaisons linéaires d'une famille fixe de vecteurs avec des coefficients statistiquement indépendants. Cette méthode a été proposée initialement pour résoudre le problème de séparation de sources en acoustique et est rapidement devenue populaire en particulier pour l'analyse de

signaux médicaux [1,2]. L'intérêt de la méthode est sa versatilité contrairement à une analyse en composantes principales (ACP) [3].

Le but de ce projet est de comprendre la modélisation sous-jacente, de la comparer à celle de l'ACP et de tester différents modèles statistiques voir d'en proposer d'autres afin d'affiner leur utilisation dans un cadre applicatif précis.

L'application visée est l'analyse de signaux d'épaisseur corticale permettant une aide au diagnostic différentiel, c'est à dire une classification des pathologies en fonction de leur incidence sur l'atrophie du cortex. Une fois les classes identifiées, les caractéristiques mises en évidence sont utilisée pour proposer un diagnostic précoce des maladies neurodégénératives.

Références

[1] Jung T.P, Makeig S., McKeown M.J., Bell A.J., Lee T.W., and Sejnowski T.J. ; Imaging Brain Dynamics Using Independent Component Analysis ; proc. Of the IEEE, vol. 89(7), 2001 http://sccn.ucsd.edu/~jung/pdf/IEEEproc01.pdf

[2] Varoquaux G. ; Sadaghiani S. ; Poline J.B. ; Thirion B. CanICA : Model-based extraction of reproducible group-level ICA patterns from fMRI time series ; fMRI data analysis workshop, MICCAI 2009

[3] Allassonnière S. ; Younes L. : A Stochastic Algorithm for Probabilistic Independent Component Analysis. S. Allassonnière, L. Younes. Annals of Applied Statistics, 2012, Vol. 6, No. 1, 125-160

2.1.2 Modèle autorégressif à régimes

Olivier Cappé cappe@telecom-paristech.fr

En économétrie (ou dans d'autres domaines...), on est parfois confronté à des phénomènes stationnaires mais présentant des périodes successives relativement distinctes. Par exemple des séries de type PIB observées sur des temps longs vont présenter des cycles assez différents qualitativement correspondant, respectivement, à des phases de croissance et de récession. (Hamilton, 1999) a proposé pour modéliser ce type de phénomène le modèle suivant :

$$(Y_t - \mu_{X_t}) = \sum_{k=1}^{p} \phi_k (Y_{t-k} - \mu_{X_{t-k}}) + \varepsilon_t$$

où (X_t) est une chaîne de Markov à valeur dans $\{0, 1\}$ (par exemple...) et μ_0, μ_1 sont des tendances correspondants par exemples au deux types de cycles (croissance/récession) mentionnés ci-dessus. (Y_t) est le processus observé, en général de type $Y_t = 100 \log(Q_t/Q_{t-1})$ où Q_t est la quantité d'intérêt. Par contre, le processus qui indique le type de cycle (X_t) est lui non directement observable (on parle également de *processus latent*). Les paramètres du modèle comprennent les coefficients autorégressifs $\phi_1, \ldots, \phi_p$, les tendances μ_0, μ_1 ainsi que la matrice de transition de la chaîne de Markov latente (X_t).

Ce modèle appartient à une catégorie plus large de modèles espace-état à régime latent, dont les *modèles de Markov cachés* constituent un autre exemple utilisé dans de nombreuses applications. Les objectifs de ce projet sont

1. de se familiariser avec les caractéristiques principales du modèle autorégressif à régimes ;
2. d'étudier et d'implémenter les équations de filtrage et de lissage non-linéaire correspondant à ce modèle ;
3. d'implémenter l'algorithme d'estimation des paramètres du modèle décrit par (Hamilton, 1990) qui repose sur le principe dit EM (pour *Expectation-Maximization*) ;
4. analyser des données analogues à celles considérées par (Hamilton, 1989).

Références

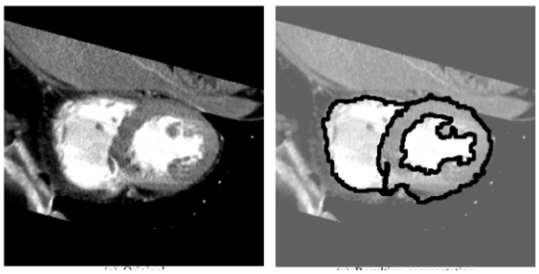
(Hamilton, 1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, Econometrica, **57**(2) :357–384.

(Hamilton, 1990) Analysis of Time Series Subject to Changes in Regime, Journal of Econometrics, **45** :39–70.

(Cappé, Moulines, Rydén, 1990) Inference in Hidden Markov Models. Springer, 2005.

2.1.3 Segmentation d'image

Stéphanie Allassonnière stephanie.allassonniere@polytechnique.edu



Un des problèmes classiques du traitement d'image est la segmentation, le découpage d'images en zones *similaires*. On souhaite étudier dans ce projet une méthode de segmentation basée sur des idées de marche aléatoire déjà utilisée dans un cadre d'imagerie médicale. Un médecin souhaite par exemple extraire la forme d'un organe dans une image médicale. Pour cela, il lui suffit de spécifier quelques points à l'intérieur de cet organe ainsi que quelques points à l'extérieur. L'objectif de ce projet est de comprendre le fonctionnement de cet algorithme, de l'implémenter et d'en proposer des améliorations.

Référence

L. Grady, "Random Walks for Image Segmentation", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 28, No. 11, pp. 1768-1783, Nov., 2006.

2.1.4 Séparation d'un mélange instantané de sources indépendantes

François Roueff roueff@cmap.polytechnique.fr

On considère le problème de la séparation d'un mélange instantané de sources indépendantes. Soient $(S_1(t))_t, \ldots, (S_K(t))_t$ K processus indépendants dont on observe un échantillon fini du mélange instantané $X(t) = AS(t)$, $t = 1, \ldots, T$, où $S(t) = [S_1(t) \ \ldots \ S_K(t)]^T$ et A est une matrice de mélange $K \times K$ inconnue. Une introduction très générale sur la séparation aveugle de sources peut être trouvée dans [1].

Dans ce projet, nous ferons l'hypothèse que les sources sont stationnaires au second ordre *par morceaux*. En particulier la matrice Cov(S(t)) est supposée constante sur des intervalles de temps $i\Delta < t \leq (i+1)\Delta$, pour $i = 0, 1, 2, \ldots$ et Δ fixé (réglé par l'utilisateur). Sur ces intervalles la matrice Cov(X(t)) est estimée empiriquement et une *diagonalisation conjointe* des matrices obtenues permet de définir un estimateur de A (à des transformations près) et donc de retrouver les sources $(S_1(t))_t, \ldots, (S_K(t))_t$ *séparées*.

Un algorithme basé sur cette idée est proposé dans [2]. Celui-ci sera appliqué à un mélange synthétisé de 2 (ou plus) signaux audio d'une dizaine de seconde (on prendra Δ correspondant à environ 20 ms).

Dans un deuxième temps, on considérera les données décrites dans [3] et disponible sur [4]. Il s'agit de signaux obtenus par des capteurs disposés à différentes

FIGURE 4 – Signaux ECG obtenus sur 8 capteurs disposés sur une femme enceinte.

positions du corps d'une femme enceinte. L'objectif est alors de séparer le signal cardiaque du foetus de celui de la mère (mélangés de façon visible sur le premier signal de la figure 4).

Références
[1] Pierre Comon, "Blind Techniques", (2010)
http://www.gipsa-lab.fr/~pierre.comon/FichiersPdf/polyD16-2006.pdf
[2] Dinh-Tuan Pham et Cardoso, J.-F., "Blind separation of instantaneous mixtures of nonstationary sources", (2001) IEEE Trans. on Signal Processing, Volume 49, Issue 9, Page(s) :1837 - 1848
[3] http://perso.telecom-paristech.fr/~roueff/edu/dataset/ecg.txt
[4] http://perso.telecom-paristech.fr/~roueff/edu/dataset/ecg.dat

2.1.5 Prédiction d'une série temporelle localement stationnaire

François Roueff roueff@cmap.polytechnique.fr

On s'intéresse à la prédiction d'une série temporelle localement stationnaire non-stationnaire. Pour la construction de prédicteurs, on utilise une modélisation de la série sous la forme d'un processus "localement stationnaire". Cette approche a été introduite notamment dans [1]. Nous nous intéresserons plus précisément à des observations $X_1, \dots, X_T$ solutions d'une équation TVAR de la forme

$$X_t = \sum_{k=1}^{p} \phi_k(t/T) X_{t-k} + \sigma(t/T)\epsilon_t \ , \qquad t = 1, \dots, T \ .$$

La suite ϵ_t est typiquement i.i.d. La non-stationarité provient du fait que les coefficients (inconnus) ϕ_k et σ sont des fonctions qui dépendent du temps t. Des estimateurs récursifs de ces coefficients sont introduits et étudiés dans le cadre de ces modèles dans [2]. Leur comportement asymptotique dépend de la variabilité des fonctions $\phi_1, \dots, \phi_k$. Plus récemment, des prédicteurs qui s'adaptent à une variabilité inconnue de ces fonctions ont été obtenus dans [3].

Le but de ce projet est d'introduire les fondements théoriques de ces méthodes de prédiction et de les appréhender à travers des expériences numériques.

Références
[1] R. Dahlhaus. "On the Kullback-Leibler information divergence of locally stationary processes." (1996) Stochastic Process. Appl., 62(1) :139–168.
[3] C. Giraud, F. Roueff, and A. Sanchez-Perez. "Aggregation of predictors for non stationary sub-linear processes and online adaptive forecasting of time varying autoregressive processes." (2015) Ann. Statist., 43(6) :2412–2450. Available at [arXiv].
[3] E. Moulines, P. Priouret, and F. Roueff. "On recursive estimation for time varying autoregressive processes." (2005) Ann. Statist., 33(6) :2610–2654. Available at [arXiv].

2.1.6 Volatilité conditionnelle localement stationnaire

François Roueff roueff@cmap.polytechnique.fr

Les modèles ARCH et GARCH ont l'avantage de permettre de modéliser avec assez peu de paramètres des séries temporelles complexes et sont pour cette raison très utilisés pour certaines séries financières, en particulier pour prédire la volatilité. Il est cependant de plus en plus admis que la prise en compte de non–stationnarités dans les séries financières est inévitable pour mettre en pratique ce type de modèles sur des données réelles de longue durée.

Pour introduire de la non–stationnarité en conservant le principe de base du modèle ARCH, il semble naturel d'introduire dans la description du modèle des coefficients variables en fonction du temps. Il reste à modéliser cette évolution temporelle pour permettre d'élaborer des méthodes d'estimation à partir de l'observation d'un échantillon. Les approches localement stationnaires [1,2] permettent de décrire cette évolution temporelle par des fonctions lisses, il s'agit d'une modélisation non–paramétrique.

Les méthodes d'estimation non–paramétriques les plus courantes reposent sur l'utilisation de noyaux de régularisation. La taille du support du noyau peut être sélectionnée par une méthode de validation croisée et des intervalles de confiance établis par des méthodes de bootstrap. La mise en oeuvre de ces méthodes dans le cadre des modèles ARCH localement stationnaire sera effectuée sur des séries financières issues de taux de change de devises et d'indices financiers tels que les rendements de l'indice Standard and Poor's 500, disponibles sur
http://finance.yahoo.com/q/hp?s=%5EGSPC

Références

[1] Piotr Fryzlewicz, Suhasini Rao and Theofanis Sapatinas, "Normalised least squares estimation in time-varying ARCH models", (2008), Annals of Statistics (36) pages 742–786.

[2] Pavel Čížek and Vladimir Spokoiny, "Varying Coefficient GARCH Models", in Handbook of financial time series (2009), pages 169-185.

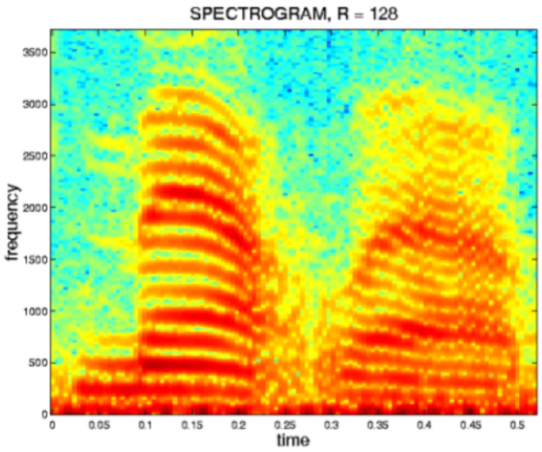
2.1.7 Vocoder de phase

Emmanuel Bacry emmanuel.bacry@polytechnique.fr

Changer la durée d'un enregistrement sonore sans en changer la tonalité ou changer la tonalité d'un enregistrement sonore sans en changer la durée sont deux besoins similaires. Le premier correspond par exemple à la nécessité lors de la diffusion des films à la télé pour compenser le passage de 24 images différents par secondes à 25 images différentes par seconde tandis que le second correspond à la nécessité de compenser le faible talent vocal de certains chanteurs. L'objectif de ce projet est d'étudier et d'implémenter un algorithme classique reposant sur des transformées de Fourier à fenêtre pour pouvoir proposer des améliorations de celui-ci.

Références

[1] J. L. Flanagan, R. M. Golden, "Phase Vocoder," Bell System Technical Journal, November 1966, 1493-1509.



2.1.8 Reconnaissance d'extraits musicaux

Emmanuel Bacry emmanuel.bacry@polytechnique.fr

Aujourd'hui il existe plusieurs logiciels permettant de retrouver le morceau de musique auquel appartient un extrait sonore. Bien entendu, ces logiciels fonctionnent à l'aide de grosses bases de données regroupant l'ensemble des morceaux potentiels. La difficulté principale vient du fait que l'extrait peut contenir de nombreuses distorsions : présence de bruits, distorsion temporelle, distorsion de hauteur... Un des logiciels les plus populaires est Shazam. Il est basé sur l'algorithme présenté ici [1]. Il consiste essentiellement à comprendre comment construire une "empreinte digitale" d'un morceau, c'est-à-dire, un ensemble restreint de features permettant de caractériser au mieux ce morceau. L'algorithme de reconnaissance consiste alors simplement à comparer (rapidement !) l'empreinte de l'extrait sonore à celle de tous les morceaux dans la base. Le but de ce projet est, dans un premier temps de comprendre le fonctionnement de cet algorithme et de le tester. Dans un second temps il faudra essayer d'identifier ses faiblesses et tenter de proposer des solutions (un exemple est donné par [2]).

Références

[1] An Industrial-Strength Audio Search Algorithm, A. Wang, Proc. 2003 ISMIR In- ternational Symposium on Music Information Retrieval, Baltimore, MD, Oct. 2003. http ://www.ee.columbia.edu/ dpwe/papers/Wang03-shazam.pdf
[2] Robust audio fingerprint extraction algorithm based on 2-D chroma, H. Wang, X. Yu, W. Wan, R. Swaminathan, International Conference on Audio, Language and Image Processing (ICALIP), 2012.

2.1.9 Détection de ruptures dans un signal

Marc Lavielle marc.lavielle@inria.fr

On considère une suite de mesures au cours du temps, ou un signal temporel, dont certaines caractéristiques peuvent changer brusquement à différents instants.

On peut penser à un signal médical comme le rythme cardiaque dont les propriétés changent suivant le type d'activité ou suivant le type de sommeil, ou l'électroencéphalogramme (EEG) dont les propriétés spectrales varient suivant l'activité électrique du cerveau. Dans d'autres domaines, on peut également penser à des indices boursiers dont la variabilité (volatilité) peut changer brutalement suivant le contexte économique, à des signaux géophysiques ou encore à des données génomiques. Lorsque ces changements brusques, ou ruptures, dans le signal se produisent à des instants inconnus, se pose alors le problème de les détecter. Il s'agit donc de localiser la position de ces ruptures (dans le temps) et éventuellement d'en déterminer le nombre, si celui-ci est inconnu. On est par conséquent confronté à un double problème statistique d'estimation de paramètres et de sélection de modèles. Des algorithmes performants ont été développés pour ce type de problème, aussi bien pour la détection en ligne de ruptures (les données arrivent en temps réel) que pour la détection hors ligne (on dispose de toute la série pour l'analyser). On s'intéressera ici tout particulièrement à des données de puces d'hybridation génomique comparative (ou CGH array) qui permettent la détection d'anomalies chromosomiques. Il s'agira dans un premier temps de bien comprendre le problème posé ainsi que la méthode de détection de ruptures décrite dans [1]. On s'appliquera dans un second temps à implémenter cette méthode et l'appliquer à ces mêmes données. Références [1] A Statistical approach for CGH microarray data analysis, Picard F., Robin S., Lavielle M., Vaisse C., Daudin J.J., BMC Bioinformatics, vol 6, n 27, 2005. http ://www.biomedcentral.com/content/pdf/1471-2105-6-27.pdf

2.2 Réseaux et graphes

Il est recommandé de suivre le cours suivant pour travailler sur les projets proposés dans cette section :

– MAP554 - Réseaux : contrôle distribué et phénomènes émergents

2.2.1 Community detection in social networks

Marc Lelarge marc.lelarge@ens.fr

In this project, we consider the problem of detecting groups of vertices with a higher-than-average density of edges connecting them. This kind of structure is called the community which is an important network property and can reveal many hidden features of the given network. Individuals belonging to the same community are probable to have properties in common. The communities in the blogspace often correspond to topics of interests. Monitoring the aggregate trends and opinions revealed by these communities provides valuable insight to a number of business 33 applications, such as marketing intelligence and competitive intelligence. Hence, identifying the communities is a fundamental step not only for discovering what makes entities come together, but also for understanding the overall structural and functional properties of a large network. Depending on the interests of the students,

we will explore connections with statistical physics, theoretical computer science, statistics or learning.

Références

[1] A tutorial on spectral clustering, Ulrike Von Luxburg, Statistics and computing, 2007
[2] Finding community structure in networks using the eigenvectors of matrices, M. E. J. Newman, Physical review E 74.3 (2006)

2.2.2 Sparse Graph Codes

Marc Lelarge marc.lelarge@ens.fr

The central problem of communication theory is to construct an encoding and a decoding system that make it possible to communicate reliably over a noisy channel. During the 1990s, remarkable progress was made towards the Shannon limit, using codes that are defined in terms of sparse random graphs, and which are decoded by a simple probability-based message-passing algorithm. In this project, the students will look at various families of sparse graph codes : three families that are excellent for error-correction : low-density parity-check codes, turbo codes, and repeataccumulate codes ; and the family of digital fountain codes, which are outstanding for erasure-correction. More recently new families of sparse graph have been introduced : polar codes, protograph codes, superposi- tion codes with connections to statistical physics, graph theory (expander graphs, random graphs) and signal processing.

Références

[1] Modern coding theory, T Richardson, R Urbanke - 2008
[2] Channel polarization : A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels, E. Arkan, IEEE Trans. Inf. Theory, vol. 55, pp.3051-3073 2009

2.3 Statistique et apprentissage

Il est recommandé de suivre l'un des cours suivants pour travailler sur les projets proposés dans cette section :

- MAP553 - The art of regression
- MAP569 - Machine learning II
- MAP566 - Statistics in action
- MAP583 - Data camp
- MAP573 - R pour les statistiques

2.3.1 Inférence et forêts aléatoires

Erwan Scornet erwan.scornet@upmc.fr erwan.scornet@polytechnique.edu

Les arbres de décision comme l'algorithme CART (Classification And Regression Trees) font partie des méthodes d'apprentissage les plus simples pour prendre en compte des corrélations entre de multiples variables. Ils sont utilisés notamment pour leur bon pouvoir prédictif et pour leur facilité d'interprétation. Cependant les arbres de décision sont connus pour être instables : une petite modification des données d'apprentissage peut entraîner d'importantes modifications dans la structure de l'arbre et ainsi conduire à un tout nouveau modèle. Ce phénomène n'est bien évidemment pas souhaitable.

Pour y remédier, des méthodes comme les forêts aléatoires, consistant à agréger de nombreux arbres de décision, ont été proposées. Les performances de ces méthodes sont très bonnes notamment dans des contextes de grande dimension, lorsque le nombre de variables d'entrée est très grand. Néanmoins, les forêts aléatoires sont difficiles à interpréter, ce qui constitue un frein à leur utilisation. Dans ce projet, nous étudierons les arbres de décision et les forêts aléatoires dans le but d'implémenter de nouvelles méthodes performantes d'un point de vue statistique et facilement interprétable. Des applications à des jeux de données réelles seront également envisagées.

2.3.2 Imputation multiple de données mixtes (numériques et catégorielles).

Julie Josse julie.josse@polytechnique.edu julie.josse@inria.fr

La problématique des données manquantes est incontournable dans la pratique statistique et pour autant la plupart des méthodes d'analyses ne peuvent pas être mises en œuvre directement à partir de données incomplètes. L'imputation multiple [1] est une méthode de référence pour faire de l'inférence en présence de données manquantes. Elle opère en trois étapes. Tout d'abord, M valeurs plausibles sont générées pour chaque valeur manquante conduisant à M tableaux imputés (complétés). Puis, la quantité d'intérêt θ et sa variance sont estimées sur chaque tableau et les estimations sont agrégées pour obtenir une estimation ponctuelle et un estimateur de la variance qui incorpore la variabilité supplémentaire due aux données manquantes et assure ainsi des taux de couverture des intervalles de confiance au niveau nominal.

Cette approche s'est démocratisée car une fois les données complétées, il est possible d'appliquer toutes les méthodes statistiques que l'on souhaite. Toutefois, les méthodes d'imputation multiple présentent encore de nombreux manques qui sont des champs de recherche actifs. En particulier, il n'existe que très peu de solutions satisfaisantes [2] pour compléter des données avec des variables mixtes, ou avec peu d'individus par rapport au nombre de variables.

Le but de ce projet est dans un premier temps de se familiariser avec la thématique des données manquantes. Dans un second temps, on se concentrera sur les techniques d'imputation multiple basées sur des décompositions en valeurs singulières pondérées qui permettent entre autres de gérer les variables catégorielles.

L'objectif est ensuite d'étendre ces travaux pour suggérer et implémenter (en R) une méthode d'imputation multiple pour des données mixtes.

La méthode proposée sera appliquée sur des données d'enquête de l'Institut National de Prévention et d'Education pour la Santé (INPES). Chaque année, l'INPES collecte des informations sur les pratiques de consommation d'alcool (nombre de verres d'alcool par mois, type de boissons consommées, etc.) d'individus adultes dans les départements français. Des informations socio-économiques et démographiques sont également disponibles. Comprendre les relations entre les variables est important pour une meilleure compréhension des pratiques et pour aider dans la définition de politiques de prévention adaptées.

Références

[1] Little, R.J.A & Rubin, D.B. (2002). *Statistical Analysis with Missing Data.* John Wiley & Sons.
[2] Murray, J. & Reiter, J. (2016). Multiple imputation of categorical and continuous via bayesian mixture models. *Journal of American Statistical Association.*
[3] Audigier, V., Husson, F. & Josse, J. (2015). MIMCA : Multiple imputation for categorical variables with multiple correspondence analysis. *Statistics and Computing.*

2.3.3 Calculs distribués et confidentialité : imputation de données médicales

Julie Josse julie.josse@polytechnique.edu julie.josse@inria.fr

Le partage de données médicales provenant de plusieurs sites (par exemple de plusieurs hôpitaux) est un enjeu majeur pour la médecine personnalisée. L'augmentation du volume d'information laisse en effet entrevoir la possibilité d'identifier plus facilement des patients aux profils similaires, d'affiner les modèles prédictifs, etc. Les tentatives d'agrégation de données ont jusqu'à présent été très peu fructueuses car de nombreux obstacles subsistent : les données ne sont pas standardisées d'un site à l'autre, il y a des problèmes de confidentialité, de propriété des données, de mises à jour de la base agrégée et aussi une difficulté à gérer une base très volumineuse. Toutefois, l'agrégation des données n'est pas indispensable et il est possible de distribuer les calculs : les données individuelles ne sont pas partagées et sont analysées sur chaque site et seuls des résumés sont mis en commun. Outre le fait de répartir les coûts de calcul, cette approche présente de nombreux avantages et est plébiscitée par la communauté. Les auteurs du projet DataShield [1] ont proposé un tel cadre pour réaliser des modèles linéaires généralisés en agrégeant des statistiques résumées anonymisées. Un premier logiciel en R [2] propose une interface complète et incorpore d'autres méthodes statistiques comme la décomposition en valeurs singulières (SVD). Il est en effet facile de réaliser une SVD globale à partir de SVD locales.

Ces projets n'ont pas encore intégré de solutions pour gerér le problème des données manquantes, problème exacerbé par l'utilisation de données provenant de plusieurs sources d'information. Ce sujet suscite de nombreux travaux que ce soit dans un cadre de méta-analyses (données agrégées) [3] ou de calculs distribués [4]. Le but de ce projet est dans un premier temps de comprendre le fonctionnement et les propriétés théoriques des techniques de SVD avec données manquantes et de complétion de matrices [5] avec contrainte de données multi-sites. Dans un second temps, il faudra intégrer la notion de confidentialité et étudier les compromis entre propriétés statistiques, calculs et confidentialités, de façon à ce que l'algorithme final conserve au mieux les propriétés théoriques de l'algorithme de base. Ce travail pourra faire l'objet d'une collaboration avec l'auteur du paquet distcomp [2], l'objectif étant d'intégrer les méthodes proposées dans ce logiciel.

Références

[1] Wolfson *et. al.* (2010). DataShield : resolving a conflict in contemporary bioscience-performing a pooled analysis of individual level data without sharing the data. *International Journal of epidemiology.*
[2] Narasimhan B., *et. al.* (2016). Software for Distributed Computation on medical Databased : A demonstration project. *Journal of Statistical Software.*
[3] Resche-Rigon, M. & White, I. (2016). Multiple imputation by chained equations for systematically and sporadically missing multilevel data.
[4] Mackey, L. Talwalkar, A & Jordan, M. (2015). Distributed matrix completion and robust factorization *Journal of Machine Learning Research.* [5] Hastie, T, *et al.* (2015). Matrix Completion and Low-Rank SVD via Fast Alternating Least Squares. *Journal of Machine Learning Research.*

2.3.4 Approche bayésienne empirique pour l'analyse de tableaux de contingence

Julie Josse julie.josse@polytechnique.edu julie.josse@inria.fr

Le modèle log-bilinéaire très utilisé pour analyser des tableaux croisés est défini de la façon suivante : $X_{ij} \sim \mathcal{P}(\mu_{ij}), \quad i = 1, \ldots, n, \ j = 1, \ldots, p,$

$$\log \mu_{ij} = \alpha_i + \beta_j + \sum_{k=1}^{K} d_k u_{ik} v_{jk}$$

L'espérance sur une échelle logarithmique se décompose donc comme en analyse de variance avec un effet ligne, un effet colonne et l'interaction est supposée avoir une structure en rang inférieur K. Pour pallier aux problèmes de surajustement très marqués dès lors que les tables de contingences contiennent de nombreux 0, de faibles comptages et que le nombre de dimension K à estimer est important, il est usuel de maximiser une vraisemblance pénalisée le plus souvent par la norme nucléaire (la somme des valeurs singulières) :

$$\hat{\mu} = \arg\min_{\mu>0} \{-\mathcal{L}(\mu) + \lambda \|\mu\|_*\},$$

Au delà du choix du paramètre de réglage (λ), il apparaît que cette pénalisation ne prend pas en compte l'hétéroscédasticité des données et par conséquent ne fournit pas une estimation satisfaisante des paramètres.

Une alternative consiste à considérer une modélisation bayésienne principalement ici pour ses propriétés de régularisation. La difficulté réside alors dans la définition des distributions *a priori* dans un cadre surparamétré : il faut par exemple définir une distribution *a priori* pour la matrice $U_{n\times K}$ qui est une matrice orthonormale. L'objet de ce projet est dans un premier temps de se familiariser avec les approches bayésiennes pour des méthodes basées sur des décompositions en valeurs singulières et d'étendre les travaux de [1] qui utilise des distributions Uniformes cas particuliers de distributions von Mises-Fisher sur la variété de Stiefel, au cadre du modèle log-bilinéaire. Différentes stratégies d'inférences seront envisagées (définition d'un MCMC) avec un intérêt particulier pour une approche bayésienne empirique dans la lignée des travaux de [2]. Les propriétés de l'estimateur rétréci seront étudiées précisément.

Les méthodes seront implémentées (en R, en utilisant BUGS, etc) et illustrées sur des tableaux carrés de données relationnelles où est enregistrée le nombre de relations entre des paires d'objets (type import-export). Selon les intérêts, il sera possible d'analyser des compétitions sportives pour classer des équipes (et prévoir les résultats de prochaines rencontres...) ou bien de considérer des données biologiques de type intéraction gènes-gènes dans le cadre d'une collaboration avec un chercheur du département des sciences du vivant de l'Imperial College of London. Ce type de données rajoute des défis : prise en compte de la symétrie, gestion des doubles 0, etc. Dans un dernier temps, on s'attachera à définir des procédures de comparaisons de modèles (par validation croisée par exemple) pour évaluer les performances des modèles développés par rapport aux très nombreux modèles concurrents [4].

Références

[1] Salmon, J. *et al.* Poisson Noise Reduction with Non-local PCA. (2013). *Journal of Mathematical Imaging and Vision.* [2] Hoff. P. (2008). Simulation of the Matrix Bingham-von Mises-Fisher Distribution, With Applications to Multivariate and Relational Data. *Journal of Computational and Graphical Statistics.*
[3] Christiansen C. & Morris C. (1997). Hierarchical poisson regression model. *Journal of the American Statistical Association*
[4] Gopalan, P., Hofman, J. & Blei, D. (2015). Scalable recommendation with hierarchical Poisson factorization. *Uncertainty in Artificial Intelligence.*

2.3.5 Analyse des données d'incidents des pompiers de Londres

Erwan Le Pennec erwan.le-pennec@polytechnique.edu

Les pompiers de Londres mettent à disposition l'ensemble des incidents où ils sont intervenus depuis le 01/01/2009. L'objectif de ce projet est d'étudier ces interventions et de proposer un algorithme de prédiction spatio-temporelle de ces

interventions. Il s'agit d'un projet typique de sciences des données où il faudra à la fois comprendre les données, en chercher des nouvelles et trouver des algorithmes permettant de bonnes prédictions.

Référence
http://data.london.gov.uk/dataset/london-fire-brigade-incident-records

2.3.6 Régression logistique, volume de données et temps de calcul

Erwan Le Pennec erwan.le-pennec@polytechnique.edu

L'objectif de ce projet est de comprendre la problématique du passage à l'échelle du *Big Data* à travers l'exemple de la régression logistique. Dans un premier temps, la régression logistique et des algorithmes de résolution numérique seront étudiés dans le cas de donnés tenant en mémoire dans un seul ordinateur et sans contrainte de temps sur les caluls. La seconde partie de l'étude sera consacré aux moyens de résoudre le même problème lorsque le volume de données augment, lorsque le temps de calcul pour la minimisation est contraint et, si le temps le permet, lorsque le temps de calcul pour la prise de décision est contraint. Le travail combinera des aspects théoriques, algorithmiques et d'implémentation

Référence
T. Hastie, R. Tibshirani and J. Friedman "The Elements of Statistical Learning" *Springer Series in Statistics.* (2009)

2.3.7 Challenge datascience ouvert

Stéphane Gaïffas stephane.gaiffas@polytechnique.edu

Le but de ce projet est de travailler sur un challenge de datascience ouvert pendant la durée des projets 3A, par exemple sur la plateforme `kaggle`. Le choix du sujet est libre, mais il faudra obtenir l'accord de l'encadrant, pour vérifier que le sujet est pertinent, réalisable dans la durée du projet, ou à l'inverse pas trop simple. Ce projet permettra d'appliquer un grand nombre de techniques d'apprentissage statistique, de comprendre leur fonctionnement pour les mettre en oeuvre correctement, d'apprendre à utiliser des librairies de machine learning et à utiliser des outils de traitement de données, dans le cadre du challenge proposé.

2.3.8 Processus de Hawkes et données MemeTracker

Stéphane Gaïffas stephane.gaiffas@polytechnique.edu

Ce projet vise à exploiter le jeu de données MemeTracker [1] qui contient des articles de presse ou entrées de blog, provenant d'environ un million de sites webs. Ce jeu de données suit les citations et phrases qui apparaissent le plus fréquemment au cours du temps sur l'ensemble de ces sites. L'objectif est alors de comprendre

comment l'information se propage, en modélisant les mécanismes qui font que certaines informations persistent, tandis que d'autres sont oubliées rapidement. Pour cela, nous proposons d'utiliser une modélisation par processus de Hawkes, qui sont une famille de processus ponctuels permettant de reproduire des effects de propagation et d'influence de sites sur les autres dans le temps. Nous mettrons à disposition des étudiants une librairie d'apprentissage statistique par processus ponctuels dévelopée au CMAP, aussi une bonne connaissance en programmation Python et C++ est recommandée pour ce projet.

Références

[1] J. Leskovec, L. Backstrom, J. Kleinberg. Meme-tracking and the Dynamics of the News Cycle. ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2009.
[2] H. Xu, M. Farajtabar, and H. Zha. Learning granger causality for hawkes processes. arXiv preprint arXiv :1602.04511, 2016.
[3] S.-H. Yang and H. Zha. Mixture of mutually exciting processes for viral diffusion. In Proceedings of the International Conference on Machine Learning, 2013.

2.3.9 L_2-Boosting et Cobra

Eric Matzner-Lober eml@uhb.fr

L'objectif de ce projet est de comprendre les méthodes de L_2-Boosting ainsi que la méthode COBRA [1] qui définit une nouvelle façon de combiner des estimateurs.

Le boosting a été à l'origine de nombreuses publications en discrimination. En régression (lorsque la variable à expliquer est continue), les publications sont moins nombreuses mais on peut citer [2] par exemple. Le principe général du boosting est de commencer avec un lisseur biaisé S_1 [2], de lisser les résidus avec le même (ou un autre) lisseur S_2, de corriger l'estimateur initial, et d'itérer. Le critère d'arrêt est en général fournit par un critère classique de choix de modèle (AIC, BIC,...). Après $k-1$ itérations, l'estimateur obtenu s'écrit de la façon suivante :

$$\begin{aligned} \hat{m}_k &= S_1 Y + S_2(I-S_1)Y + \cdots + S_k(I-S_{k-1})\cdots(I-S_1)Y \\ &= [I-(I-S_k)(I-S_{k-1})\cdots(I-S_1)]Y. \end{aligned}$$

Si le même estimateur est utilisé à chaque étape on obtient une forme simplifiée

$$\hat{m}_k = [I-(I-S)^k]Y.$$

Ces estimateurs ont été étudiés dans [3] par exemple.

Dans [1], une nouvelle méthode d'agrégation est présentée tant du point de vue théorique que pratique. Dans toutes les méthodes d'agrégation, il est intéressant d'avoir des estimateurs admettant des comportements différents. L'objectif de ce projet est d'analyser numériquement l'apport de méthode du type L_2-Boosting (additif [2] ou multivarié [3]) dans le cadre de COBRA.

Références
[1] Biau, G., Fischer, A., Guedj, A. et Malley, J. COBRA : A combined regression strategy, submitted
[2] Bühlmann, P. et YU, B. Boosting with the l_2 loss : regression and classification, *JASA*, 324-339, 2003.
[3] Cornillon, P-A. et Hengartner, N. et Matzner-Løber, E. Recursive bias estimation for multivariate regression smoothers, *ESAIM*, 2013.

2.3.10 Modélisation markovienne de variables météo

Denis Talay (CMAP et Inria) denis.talay@inria.fr
Mireille Bossy (Inria) mireille.bossy@inria.fr

Les études climatologiques utilisent plusieurs échelles temporelles et spatiales, allant de millénaires et de l'échelle globale de la planète quand il s'agit des bouleversements climatiques majeurs du passé et à venir, à l'heure et jusqu'au km^2 quand il s'agit de prédire le temps qu'il fera demain.
Les modèles météorologiques statistiques ont pour objet d'aboutir à une description statistique des événements climatiques, à la modélisation et la simulation de phénomènes récurrents ou/et extrêmes, ainsi qu'à l'analyse de risques divers (écologiques, financiers, économiques, énergétiques, etc.) liés aux événements extrêmes (grands froids pendant de longues périodes, ensoleillement pour le tourisme et l'estimation de production d'énergie solaire, par exemple).
Ce projet aura pour objet de simuler et calibrer un modèle markovien pour certaines variables climatiques et météorologiques. Compte tenu de la fréquence assez grossière des données généralement disponibles (de l'ordre d'un ou deux relevés par jour), il est naturel de privilégier une modélisation en temps discret. Comme le modèle étudié prend notamment en compte les "régimes de temps" sur l'Europe, des couplages entre différentes variables météorologiques, et l'existence de cycles (diurnes, saisonniers), il est bien plus complexe que les modèles classiques de type ARMA et demande des techniques de simulation et de calibration spécifiques.
Des données météorologiques réelles seront disponibles et permettront de vérifier la validité du modèle et des procédures d'estimation étudiées.
Ce projet permettra d'étendre les connaissances en analyse, simulation, et calibration statistique de processus de Markov ergodiques à temps discret et d'apprendre quelques rudiments de la théorie ergodique des processus.

Références
[1] D. Dacunha-Castelle et M. Duflo. Probability and Statistics. Vol. II. Springer-Verlag, 1986
[2] S.P.Meyn et R.L.Tweedie. Markov Chains and Stochastic Stability. Springer-Verlag, 1993

3 Optimisation et recherche opérationnelle

Enseignants référents : Xavier Allamigeon xavier.allamigeon@inria.fr et Stéphane Gaubert Stephane.Gaubert@inria.fr

3.1 Etude de modélisation du coût de retard avion et modèle d'optimisation de la ponctualité avion

Marine Le Touzé (maletouze@airfrance.fr), Air France
Blaise-Raphael Brigaud, Air France

Dans un contexte concurrentiel de plus en plus agressif, la ponctualité avion est un enjeu primordial pour les compagnies aériennes :

- enjeu commercial et satisfaction client ;
- enjeu économique et rentabilité.

Chaque minute de retard induit des coûts additionnels aux opérations de la compagnie aérienne.

Le travail sera construit en deux axes :

- construire un modèle de coût du retard avion (utilisation des infrastructures aéroport, immobilisation équipage, répercussions commerciales, ...)
- proposer et tester plusieurs algorithmes d'optimisation prenant en compte la robustesse opérationnelle, pour la construction des rotations avion.

3.2 Etude des stratégies et évaluation de la performance pour le Revenue Management AF/KL

Wail Benfatma, Air France
Michael Chalamel (michalamel@airfrance.fr), Air France

Au sein d'Air France – KLM, le Revenue Management est l'optimisation de la recette par la gestion de la disponibilité des tarifs en fonction des capacités restantes et de prévisions de demande. En pratique, la problématique réellement traitée par les transporteurs aériens comme Air France KLM présente de nombreux aspects de complexité : grand nombre de vols, structure du réseau de transport, multiples tarifs, volatilité de la demande, ...

L'enjeu de ce projet sera la mise en place d'un modèle permettant d'estimer a posteriori la performance du Revenue Management. L'objectif est de déterminer les stratégies optimales et de dégager des indicateurs pour améliorer les stratégies de gestion des vols.

3.3 Optimisation de la palettisation et placement d'un avion cargo

Ferran Garcia (fegarciarigau@airfrance.fr), Air France
Magdalena Kociolek (makociolek@airfrance.fr), Air France

Dans le monde du air cargo, la marchandise peut être transportée dans les avions tout cargo ou passagers. Elle est rangée dans différentes palettes, dont le placement et dimensions varient d'un avion à l'autre. Il arrive que les marchandises réservées pour un vol excédent sa capacité. Le cas échéant, certaines marchandises doivent être débarquées et placées sur un autre vol.

Le but sera dans un premier temps de trouver une méthode de palettisation et placement efficace, et dans un deuxième temps de déterminer quelles marchandises il faut débarquer si la capacité est insuffisante afin de maximiser le profit total du vol.

3.4 Estimation de revenu d'un programme de vol

Solène Richard, Air France
Anne-Laure Cebile, Air France

La création du programme de vol d'une compagnie aérienne est un processus complexe qui vise à trouver une combinaison de vols attractifs pour la clientèle.

Pour le faire efficacement, le choix de l'agencement horaire des vols est une étape cruciale car il va déterminer l'attractivité de l'offre Air France par rapport à celle de la concurrence et donc la demande que celle-ci va susciter du côté des passagers. La part de la demande que la compagnie peut effectivement capter est limitée par les capacités des vols dont le trajet est constitué.

L'estimation de la répartition des passagers sur le réseau, puis du revenu qu'ils peuvent générer, doit donc prendre en compte la demande sur chacun des trajets (local et connecting), ainsi que les contraintes de capacité de chacun des vols.

Afin d'estimer le revenu maximal que pourrait générer un programme de vol donné, on pourra modéliser le réseau de la compagnie et estimer la répartition des passagers sur les différents vols et trajets existants, sous contrainte de capacité et de demande.

3.5 "Acteur local" dans les réseaux électriques du futur : prise en compte du risque dans un modèle Markov Decision Process (MDP)

Olivier Beaude (olivier.beaude@edf.fr), EDF R&D, Dpt OSIRIS

Dans les réseaux d'électricité du futur (plus connus sous leur appellation anglophone "Smart Grids"), le problème de l'acteur local apparaît dans de nombreuses applications en plein développement. Cet acteur générique – souvent petit

et géographiquement localisé – gère un ensemble de consommateurs électriques, dont certaines sont flexibles (elles peuvent être modifiées, décalées, etc), en disposant d'un système de production local intermittent (panneau photovoltaïque par exemple, dont la production est incertaine) et éventuellement d'un système de stockage. L'objectif de cet acteur est de satisfaire les besoins de ses consommateurs locaux en minimisant sa facture d'électricité. Cette facture est calculée sur la base de prix variables, qui peuvent être directement les prix de marchés Spot d'électricité. En adoptant un modèle markovien pour les incertitudes, ce problème peut être traité avec des chaînes de Markov contrôlées (Markov Decision Processes). Cette approche est maintenant classique dans la littérature. La nouveauté à tester dans ce projet serait d'intégrer une notion de risque dans ce problème : risque lié à la consommation flexible (un véhicule électrique peut quitter plus tôt que prévu son lieu de charge), risque lié aux prix de l'électricité (les prévisions sur les prix de l'électricité sont rarement parfaites !)...Cette approche a été proposée dans un article très récent [1] pour la charge des véhicules électriques, nous proposons de l'analyser théoriquement et de la mettre en pratique en simulation sur des cas pratiques étudiés à la R&D d'EDF (charge intelligente d'un véhicule électrique, contrôle d'usages électriques domestiques...). Les simulations se feront sous le langage Python.

Références

[1] D.R. Jiang and W. B. Powell, Optimal policies for risk-averse electric vehicle charging with spot purchases, soumis en mai 2016.

3.6 Modélisation des prix fondamentaux du minerai de phosphate

Riadh Zorgati (riadh.zorgati@edf.fr), EDF R&D, Dpt OSIRIS

Le minerai de phosphate permet de produire des engrais qui jouent un rôle-clé dans la production agricole. En raison d'une prévision de hausse de la demande en engrais, liée à une augmentation de la population mondiale de près de 50% d'ici à l'horizon 2050, ce minerai est d'importance stratégique et a engendré un fort intérêt se traduisant par des études sur sa disponibilité dans le futur et sur son prix. Le prix de ce minerai est réputé très volatile et pourrait résulter à la fois de considérations géologiques et économiques, en particulier d'équilibre offre-demande, des jeux d'acteurs sur les marchés et des incertitudes du problème : réserves, épuisement des mines existantes, développement des pays émergeants, etc.

Le projet a pour objectif de développer un modèle simplifié d'estimation du prix du phosphate fondé sur un équilibre offre-demande. En fonction de l'avancement des travaux, le modèle pourrait être enrichi, par exemple par des considérations sur la structure des marchés.

Les principales étapes du projet sont :

– une analyse bibliographique limitée (documents fournis)

– une modélisation du problème (dans un premier temps un programme linéaire)
– une collecte des données nécessaires au modèle
– un codage en langage python
– une analyse des premiers résultats obtenus avec comparaison des prix observés, permettant de guider vers quels points il faudrait travailler pour enrichir le modèle
– un enrichissement du modèle selon l'analyse précédente (théorie des jeux, épuisabilité de la ressource, "mechanisme design", etc.)

3.7 Localisation optimisée de dépôts avec prise en compte du calcul de tournées

Bayram Kaddour (bayram.kaddour@edf.fr), EDF R&D, Dpt OSIRIS

Un dépôt est un centre à partir duquel des véhicules partent pour effectuer des tournées pour desservir des points de livraison dans un pas de temps avant d'y retourner. Le problème consiste à trouver l'emplacement optimal de ces dépôts afin de couvrir l'activité de livraison prévue. Les choix des dépôts seront pris au cours de la période étudiée pour prendre en compte une évolution de l'activité sur la zone géographique concernée. Dans le cas où la localisation des dépôts est fixée, le problème devient un problème classique de calcul de tournées de véhicules. Ces deux problèmes sont donc corrélés et l'objectif du projet consiste à modéliser, implémenter et tester une approche qui permet la prise en compte du calcul de tournées dans le problème de localisation de dépôts sans passer par des simplifications grossières.

3.8 Prise en compte des accès wifi dans l'optimisation de la sélection des accès radios cellulaires

Mustapha Bouhtou (mustapha.bouhtou@orange.com), Orange Labs Recherche

Les terminaux mobiles actuels (smartphones) sont équipés d'interfaces leur permettant d'accéder à plusieurs types de réseaux d'accès disponibles (2G/3G/LTE/WIFI, ...). Les usagers peuvent donc se connecter au réseau d'accès leur offrant la meilleure connectivité dans la limite de leurs contrats. Cela pourrait à certains endroits et certaines périodes induire des surcharges importantes sur un réseau d'accès particulier avec une dégradation de la qualité de service pour les clients. Dans ce contexte se pose alors la question de comment bien répartir de façon dynamique le trafic des clients d'un opérateur entre ces différents réseaux d'accès disponible. L'objectif étant de permettre à chaque terminal de pouvoir automatiquement se connecter au bon réseau au bon moment et au bon endroit. Les préférences à la fois des utilisateurs et de l'opérateur sont à considérer.

Dans ce travail on essaiera de modéliser et simuler la dynamique d'affectation des flux de trafic en temps réel aux différents réseaux d'accès, d'étudier l'optimisation du système global et de proposer des algorithmes pour le résoudre. Le cas de plusieurs services, avec plusieurs profils de clients sera abordé. Les contraintes sur la mobilité des clients devront aussi être traitées.

3.9 La vidéo streaming et le control des périodes de connexion (modes on/off) des smartphones : optimisation conjointe de l'énergie et de la qualité d'expérience

Mustapha Bouhtou (mustapha.bouhtou@orange.com), Orange Labs Recherche

Aujourd'hui les terminaux mobiles sont de plus performants en termes de puissance de calcul, de mémoire, de taille et de la qualité de l'écran. Ces performances ont aussi fortement stimulé la demande de services multimédias notamment le service de vidéo streaming. Les clients sont aussi de plus en plus exigeants quant à la qualité rendue et perçue sur ce type de service. Optimiser la qualité d'expérience (QoE) des clients sur les services de vidéo streaming est pour cela un enjeu important pour les opérateurs.

Le problème de la modélisation de la QoE a déjà été abordé dans la littérature. Dans ce projet on s'intéressera particulièrement à une modélisation basée sur des processus stochastiques (chaines de Markov) pour prendre en compte la dynamique des arrivées des clients sur le réseau mais aussi celle des flux de trafics que génère leurs demandes en services de streaming. On étudiera comment introduire un critère d'optimisation dans cette modélisation et on traitera ensuite comment le résoudre. Des tests numériques seront à effectuer et les résultats devront être comparés à ceux de la simulation.

3.10 Routage d'une matrice de trafic multi-horaire

Eric Gourdin (eric.gourdin@orange.com), Orange Labs OMN/NMP/TRM

Contexte Télécom

Le nombre toujours croissant de services offerts aux usagers de l'Internet fixe et mobile, particuliers ou entreprise, génère des volumes de plus en plus important de trafic qu'il faut souvent écouler dans les grands réseaux d'interconnexion internationaux. La gestion opérationnelle de ces grands réseaux devient donc un enjeux considérable pour les opérateurs de ces réseaux, d'autant que la compétition entre opérateurs les oblige à veiller à une utilisation aussi efficace et parcimonieuse que possible des ressources du réseaux. En parallèle, les incidents et les opérations de maintenance dans le cœur du réseaux ont un impact considérable sur la qualité des services offerts aux clients. Les principes de gestion de ces réseaux consistent donc souvent à définir des plans de routage efficaces, stables et robustes et surveiller

les évolutions de la demande pour décider, de temps en temps, soit d'adapter le réseau lui même (ajouter ou augmenter les capacités, ajouter des liens, ...), soit de modifier le plan de routage.

Même si des évolutions sont proposées régulièrement pour rendre plus flexibles les protocoles de routage dans les réseaux IP, les principes de routage utilisés en pratique restent essentiellement basés sur des notions de mono-routage (un chemin unique pour chaque demande, c'est-à-dire, pour chaque paire origine/destination) et de routage au plus court chemin (selon des valeurs sur les arcs définies par l'administrateur réseau et qui constitue la "métrique administrative"). Plus précisément, les protocoles "natifs" (OSPF, IS-IS, ...) routent les demandes sur des plus court chemins et des évolutions plus récentes (MPLS, Segment Routing, ...) permettent de définir des chemins indépendamment des valeurs de métrique. Pour résumer, on peut dire qu'une grande majorité du trafic est routé selon les protocoles classiques (plus court chemins) mais que quelques chemins additionels peuvent également être définis pour desservir quelques demandes particulières.

Pour gérer leur réseau, les opérationnels disposent d'un grand nombre de données, à savoir, d'une part des mesures de charge quasiment en temps réel sur tous les liens du réseau, et d'autre part, des mesures ou estimations relativement précises, mais sur une échelle de temps un peu plus longue (quelques minutes, heure, ...), de la demande totale écoulée de bout-en-bout, pour chaque paire origine/destination. Ces dernières données sont généralement rassemblées dans ce qu'on appelle une matrice de trafic.

Si les problèmes d'optimisation de réseaux sont relativement clairs et faciles à poser lorsqu'on dispose d'une unique matrice de trafic (représentant, par exemple, une projection dans le futur de l'heure chargée), la question est beaucoup plus difficile lorsqu'on dispose d'une série de matrices de trafic et que l'on souhaite utiliser directement les informations qu'elles contiennent dans un modèle d'optimisation. Il existe par exemple des modèles dit "multi-hour network design" [1,2,3] où le réseau doit être dimensionné pour permettre d'écouler chacune des matrices de trafic (pire cas) ou bien où l'on s'autorise à modifier légèrement le réseau à chaque nouveau pas de temps.

Analyse conjointe des matrice de trafic et du réseau

Dans ce projet, on s'intéresse à une autre façon d'utiliser une matrice de trafic multi-période. L'objectif est de définir une stratégie globale de routage adaptée à un réseau existant, utilisant les informations des matrices de trafic et "facile" à mettre en œuvre d'un point de vue opérationnel. On veut donc définir l'ensemble des routes qui doivent être mises en œuvre dans le réseau avec un objectif global de "bonne utilisation des ressources", qu'on traduira par : minimiser la charge du lien le plus chargé.

L'autre importante contrainte opérationnelle consiste à construire un plan de routage global qui change le moins possible et le moins souvent possible. On pourra aussi consiérer qu'on part d'un plan de routage initial que l'on veut modifier le

moins possible. L'objectif du projet est d'étudier et de mettre en œuvre des modèles d'optimisation permettant de répondre à cet objectif général.

Travail attendu

On se donne un réseau modélisé par un graphe orienté (ou bi-orienté) avec des capacités sur les arcs et une matrice de trafic multi-horaire, qui donne le débit à écouler dans le graphe pour chaque paire o/d (paire de nœuds du graphe, o = origine, d = destination) et chaque tranche horaire (h = 1,2,. . .,24).

On partira d'un plan de routage global unique à construire, c'est-à-dire, l'ensemble des chemins, un chemin par demande (paire o/d) qui minimise le critère (charge du lien le plus chargé) sur l'ensemble des périodes (pire cas).

Puis, on cherchera comment améliorer le critère si l'on s'autorise à changer de routage une fois, ou deux fois (aux tranches horaires à déterminer).

On considérera ensuite le probème où l'on peut modifier le routage à n'importe quelle tranche horaire, mais uniquement sur un sous-ensemble de demandes (très) limité.

D'autres variantes pourraient également être envisagées (combinaison de mono-routage et routage au plus court chemins, prise en compte de pannes, évolution à plus long terme de la matrice de trafic, etc)

Le codage des modèles se fera de préférence en Python avec la bibliothèque Pyomo/Jupyter pour l'appel aux solveurs (Cplex, cbc, glpk, . . .) et la bibliothèque Networkx pour la manipulation des graphes.

Références

[1] W. Ben-Ameur. Multi-hour design of survivable classical ip networks. *International Journal of Communication Systems*, 15 :553–572, 2002.

[2] Deep Medhi and David Tipper. Some approaches to solving a multihour broadband network capacity design problem with single-path routing. *Telecommunication Systems*, 13 :269–291, 2000.

[3] J.M. Hattingh S.E. Terblanche, R. Wess´aly. Solutions strategies for the multihour network design problem. In *INOC 2007*, 2007.

3.11 Conception d'un CDN robuste (Content Delivery Network)

Eric Gourdin (eric.gourdin@orange.com), Orange Labs OMN/NMP/TRM

Contexte Télécom

La demande toujours plus importante en contenus vidéos délinéarisés (UGC, Web TV, replay, . . .) oblige les acteurs de l'Internet à optimiser le placement et la réplication des contenus dans le réseau de distribution. Les grands acteurs de l'Internet multiplient leurs sites de stockage de contenus et proposent aux opérateurs et aux fournisseurs d'accès des offres permettant de stocker les contenus

les plus populaires au plus proche des clients finaux. L'élèment déterminant dans la conception d'un CDN réside dans le compromis entre le coût des serveurs de stockage additionels et le gain en bande passante observé dans le cœur de réseau. Le gain est d'autant plus important que les serveurs de stockage sont à même de répondre souvent aux requêtes des clients. Un des équipements les plus utilisés dans les CDNs est le cache.

Un cache est un serveur de stockage particulier qui adapte dynamiquement les contenus qu'il stocke en fonction des arrivées de requêtes. Supposons que tous les contenus disponibles soient initialement hébergés dans un serveur central. A chaque arrivée de requêtes (émanant de clients "proches"), le cache va servir directement le client s'il a le contenu. S'il n'a pas le contenu, il prolonge la requête vers le serveur central qui va le lui transmettre. A la reception du contenu, le cache stocke le contenu et le transmet au client. A terme, si la capacité du cache était infinie, il finirait par stocker l'ensemble des contenus (ou catalogue). Comme la capacité des caches est généralement limitée par rapport à la taille du catalogue (qui peut être très grand, par ex. toutes les vidéos Youtube), le cache doit, de temps en temps, décider de supprimer certains contenus. Il existe plusieurs politiques classiques pour décider des contenus à supprimer : par exemple, LRU (Least Recently Used) consiste à supprimer le contenu dont la date de la dernière requête est la plus ancienne, FIFO (First In First Out) consiste à supprimer le contenu qui à été inséré le premier, etc. L'efficacité d'un cache, mesurée par son Hit Ratio (proportion des requêtes qu'il est capable de satisfaire), dépend essentiellement du choix de la politique de replacement et de la distribution des popularités des contenus. La popularité d'un contenu est la fréquence (sur une période de temps à définir) avec laquelle des requêtes pour ce contenu sont émises. Il a été observé que, pour la plupart des types de contenus, la popularité suit une distribution de type Zipf, où très peu de contenus sont très populaires, et la très grande majorité des contenus est très peu populaire.

Dans un problème de conception de CDN, la décision d'installer des caches est prise de manière statique alors que son efficacité dépendra d'une distribution de popularités future, et donc inconnue et de la dynamique des arrivées de requêtes. Une façon simplifiée de prendre en compte cette dynamique a été proposées dans [4].

Un modèle simplifié

Dans [4], on doit installer des caches dans un réseau arborescent et l'on suppose que les requêtes émanent des feuilles de l'arbre. On suppose également que la distribution de popularités des contenus reste relativement stable dans le temps et peut donc être utilisée pour dimensionner le CDN. Pour simplifier le problème, on suppose enfin que le catalogue est subdivisé en un nombre restreint de classes (typiquement, une dizaine) construites en considérant les contenus par popularités décroissantes (la classe 1 contient donc les contenus les plus populaires). La capacité des caches est exprimée en fonction du nombre de classes qu'il peut stocker et la dynamique de fonctionnement des caches est modélisée de manière approchée

en imposant aux caches installés de stocker les classes les plus populaires dont il voit passer les requêtes. En particulier, si un autre cache, par exemple de capacité 2, a été installé plus bas dans le réseau (plus proche des clients), alors le cache suivant pourra stocker les classes 3, 4 ou encore moins populaires. Un modèle MIP (Mixed Integer Program) est proposé et utilisé pour résoudre des instances de taille raisonnable.

L'objectif du projet est d'étudier deux extensions au modèle [4].

Extension 1 : routage des demandes dans un graphe quelconque

Routage des demandes Dans [4], le graphe modélisant le réseau est un arbre et, par conséquent, les routages des requêtes vers le serveur central et/ou les caches sont fixés. On va s'intéresser au problème plus général où il existe plusieurs chemins potentiels pour qu'un client accède aux serveurs. Le choix du chemin fera partie de l'optimisation et l'on utilisera un modèle de multiflot pour gérer le choix du ou des chemins à utiliser.

Travail attendu Le modèle de localisation de cache devra être étendu de manière à permettre un choix optimal de routes entre les nœuds de demandes, les caches et le serveur central. Un coût d'utilisation des arcs proportionnel à la charge devrait conduire à router vers les caches les plus proches. Des contraintes additionnelles sur la capacité des arcs pourraient forcer le choix de chemins un peu plus longs. Le cache permettant de satisfaire les requêtes d'une classe n'étant pas connu à l'avance, il faudra paramétrer le modèle de flot en fonction des nœuds potentiels d'arrivée des demandes.

Le codage des modèles se fera de préférence en Python avec la bibliothèque Pyomo/Jupyter pour l'appel aux solveurs (Cplex, cbc, glpk, ...) et la bibliothèque Networkx pour la manipulation des graphes.

Extension 2 : prise en compte de l'incertitude sur la demande

Programmation robuste La programmation stochastique permet d'introduire des éléments incertains au moyen de variables aléatoires dans des modèles d'optimisation [1]. Même si ces modèles sont très précis dans leur façon de prendre en compte l'incertain, ils sont généralement très difficiles à résoudre. C'est pourquoi, il y a une quinzaine d'années, les approches dites de programmation robuste ont été proposées : dans ce type d'approche, l'incertitude est modélisée de manière déterministe en construisant des ensembles où les variables incertaines peuvent prendre leurs valeurs et en intégrant dans la fonction objectif l'aversion au risque que l'on est prêt à accepter [2,3]. Pour que les problèmes puissent être résolus en pratique, les ensembles d'incertitude généralement considérés sont des structures relativement simples (intervalles, ellipsoides, polyèdres, ...). Plus récémment, une approche "multi-bandes" a été proposée pour modéliser plus finement des formes d'ensembles d'incertitude plus complexes, avec des bandes successives (voir www.dis.uniroma1.it/~fdag/multiband.html). Parallèlement, des modèles de

robustesse à deux étapes ont également été proposés [5] : dans ces modèles, des décisions stratégiques sont prises au premier niveau (par exemple, où déployer des caches) et un recours peut être utilisé au deuxième niveau lorsqu'on observe la réalisation des variables de deuxième niveau (par exemple, distribution des popularités).

Travail attendu L'objectif du projet est d'étudier quelles approches sont les plus adaptées pour modéliser l'incertitude sur la distribution de popularités. En particulier, on mettra en œuvre des modèles robustes multi-bandes et à deux étapes pour tenter de résoudre des instances de problèmes. Les résultats obtenus par ces modèles d'optimisation pourront être validés (ou invalidés) au travers de simulations.

Le codage des modèles se fera de préférence en Python avec la bibliothèque Pyomo/Jupyter pour l'appel aux solveurs (Cplex, cbc, glpk, ...) et la bibliothèque Networkx pour la manipulation des graphes.

Références

[1] John R. Birge and François Louveaux. *Introduction to Stochastic Programming*. Springer-Verlag, New York, NY, USA, 1997.
[2] C. Caramani D. Bertsimas, D. Brown. Theory and applications of robust optimization. *SIAM Review*, 53(3) :464–501, 2011.
[3] D. Brown D. Bertsimas. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6) :1483–1495, 2009.
[4] E. Gourdin and P. Bauguion. Optimal hierarchical deployment of caches for video streaming. In *Network of the Future (NOF), 2015 6th International Conference on the*, pages 1–5, Sept 2015.
[5] Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *OR Letters*, 41(5) :457–461, 2013.

3.12 Optimisation stochastique d'une batterie à partir de plusieurs sources intermittentes

Karim El Alami karim.elalami@elum-energy.com

Cyril Colin cyril.colin@elum-energy.com

L'objectif de ce sujet est de réaliser dans un premier temps une optimisation entre plusieurs sources d'énergie (panneaux photovoltaïques, réseau, générateur diesel), une batterie et une consommation électrique en fonction de contraintes liées à la facture énergétique. Puis dans un second temps, une amélioration de la robustesse des algorithmes d'optimisation sera réalisée en s'appuyant sur le fait que l'optimisation se base sur une prédiction de production photovoltaïque intermittente et une prédiction de consommation incertaine. Les algorithmes que nous souhaitons utiliser reposent sur le contrôle prédictif et le contrôle stochastique.

Références
[1] M. Encina , *Microgrid Management - Battery Aging Cost*, Coppetti's Model Implementation 2015
[2] F. Katirei, *Microgrid Management*, IEEE Power and Energy Magazine 2008
[3] P. Haessig, *Computing an Optimal Control Policy for an Energy Storage*, PROC. of the 6th eur conf on python in science (euroscipy 2013)

3.13 Optimisation de centre d'appel

Xavier Allamigeon (xavier.allamigeon@inria.fr), INRIA – Ecole polytechnique
Stéphane Gaubert (stephane.gaubert@inria.fr), INRIA – Ecole polytechnique
Avec la participation de Pascale Bendotti et Agnès Bialecki (EDF R&D).

Contexte Ce projet a pour origine un problème de gestion de compétences des conseillers clientèle nécessaires pour répondre aux demandes des clients dans les centres d'appel d'EDF. Ces demandes arrivent par diverses canaux (appels téléphoniques, email, courrier) et concernent différentes activités. Les activités traitées se distinguent principalement par leur caractère front-office ou back-office. Dans les activités front-office, le conseiller clientèle prend en charge un client d'EDF au téléphone et instruit sa requête en temps réel. Les activités back-office sont constituées d'activités où le conseiller clientèle n'est pas en relation immédiate avec le client, mais elles nécessitent un traitement dans une durée limitée (par exemple, réponse par email à un client sous 24 ou 48 heures). Les conseillers proviennent de deux types de ressources :

- internes : elles doivent être gérées par EDF, en prenant en compte leur disponibilité (congés, formation, RTT, maladie) et une limite sur la nature et le nombre d'activités différentes qu'elles peuvent traiter. A noter que le volume de ressources internes est donné mais fluctuant et contraint par activité.
- externes : les contrats de prestation se font en termes de service à rendre et non pas de moyens à mettre en œuvre. Les ressources correspondantes n'étant pas gérées par EDF, les fluctuations des ressources externes (disponibilité, limite sur la nature et le nombre activités) ne sont pas à prendre en compte.

L'idée est de déterminer le volume de ressources internes-externes, avec les compétences associées, nécessaire par activité pour traiter les demandes.

Pour le problème de gestion des compétences, il apparaît nécessaire d'estimer la capacité de traitement des conseillers clientèles à une échelle de temps courte afin de tenir compte de la fluctuation des effectifs des ressources internes, et des demandes (écart de prévision, saisonnalité). L'objectif est donc d'évaluer la performance d'un système faisant apparaître des contraintes temps réel sur les ressources et sur le temps de traitement des tâches pour satisfaire un flux de demande par activité.

Programme de travail. On se propose de modéliser le centre d'appel d'EDF et analyser ses performances en vue d'optimiser son dimensionnement. On utilisera

pour cela des méthodes issues de l'algèbre et la géométrie tropicales. Les questions d'évaluation de performance furent à l'origine même du développement de techniques de nature tropicale, avec les travaux de Cohen, Quadrat et Viot puis de Baccelli et Olsder (cf. [1,3]) dans les années 80. Ces travaux furent prolongés par le développement du *network calculus* fournissant des estimations analytiques de qualité de service [2,4]. Cependant, les systèmes considérés étaient limités à des classes bien particulières (par exemple, les graphes d'évènements). Ce type d'approche a été récemment revisité pour l'étendre à des systèmes mêlant synchronisation et concurrence, plus représentatifs des applications réelles [5]. Grâce à cela, on a analysé les performances d'un centre de réception d'appels d'urgence prochainement mis en place par la Préfecture de police de Paris et la Brigade de sapeurs-pompiers de Paris. La nouveauté du problème posé par le centre d'appels d'EDF réside dans l'analyse des régimes transitoires (absorption des pics de charge), et le traitement par les mêmes personnels d'activités avec des constantes de temps différentes. Les méthodes décrites précedemment devraient pouvoir s'appliquer au prix d'un important travail d'adaptation.

Le projet comprendra donc un premier voler de modélisation du centre d'appel à l'aide d'un formalisme comme les réseaux de Petri. Il s'agira ensuite d'implémenter un simulateur afin de confronter la modélisation aux observations réelles. Enfin, si le temps le permet, on s'intéressera à l'analyse mathématique du modèle à l'aide des techniques sus-citées.

Références

[1] F. Baccelli, G. Cohen, G.J. Olsder, and J.P. Quadrat. *Synchronization and Linearity*. Wiley, 1992.

[2] Cheng-Shang Chang. *Performance Guarantees in Communication Networks*. Springer-Verlag, London, UK, UK, 2000.

[3] G. Heidergott, G. J. Olsder, and J. van der Woude. *Max Plus at work*. Princeton University Press, 2006.

[4] Jean-Yves Le Boudec and Patrick Thiran. *Network Calculus : A Theory of Deterministic Queuing Systems for the Internet*. Springer-Verlag, Berlin, Heidelberg, 2001.

[5] X. Allamigeon, V. Bœuf, and S. Gaubert. Performance evaluation of an emergency call center : tropical polynomial systems applied to timed Petri nets. In *FORMATS'15*, volume 9268 of *Lecture Notes in Computer Science*. Springer, 2015.

3.14 Optimisation temps réel de la conduite d'un centre d'appel 17-18-112

Xavier Allamigeon (Xavier.Allamigeon@inria.fr), INRIA – Ecole polytechnique
Stéphane Gaubert (Stephane.Gaubert@inria.fr), INRIA – Ecole polytechnique

La préfecture de police a proposé une réforme de la procédure de traitement des appels d'urgence à Paris et en petite couronne. La nouvelle procédure, en cours de déploiement, permet de traiter dans un cadre unifié les appels concernant les

numéros 17 (police), 18 (pompiers), ou 112 (numéro européen indifférencié). Dans cette nouvelle procédure, un opérateur de premier niveau filtre les appels, assure le cas écheant une mission de conseil, et transmet à un second niveau les appels qui nécessitent une instruction par un opérateur spécialisé (policier ou pompier).

La question de l'évaluation de performance et du dimensionnement d'un centre d'appel de ce type a été abordée depuis deux ans dans le cadre d'enseignements d'approfondissement et de projets de 3A à l'École polytechnique, proposés pour répondre à une demande de la Préfecture de police, avec l'appui de la Brigade de sapeurs-pompiers de Paris, et encadré par des enseignants et chercheurs de l'École. Ces travaux ont permis de développer un modèle détaillé du centre d'appel, à base de réseau de Petri, rendant compte des procédures de traitement, des règles de priorité, ainsi que de l'aspect bi-métier (police et secours) et de la possibilité d'avoir des chaînes de traitement différenciées selon l'origine géographique de l'appel. Ce modèle a été calibré sur des données réelles, incluant à la fois des journées ordinaires et des journées avec des "coups de béliers" (par exemple, orage violent assorti d'inondations conduisant à un grand nombre d'appels 18). Cela a permis d'évaluer, par des calculs analytiques ou des simulations, des indicateurs de performance (temps d'attente, taux d'appels raccrochés). Cela a aussi permis de mettre en évidence différentes phases de congestion potentielles, qui peuvent apparaître selon le débit d'appel entrants, les ressources disponibles, et les règles de traitement des appels. L'analyse de ces différente phases peut aider à dimensionner le système.

On se propose cette année de s'intéresser à la problématique de l'optimisation de la conduite d'un centre d'appel de type 17-18-112. Un tel centre peut être contrôlé de plusieurs manières. Il est possible par exemple de rappeler rapidement des opérateurs de type pompiers pour répondre à un afflux massif d'appels. Il est aussi possible de recommander aux opérateurs d'abréger ou de différer la réponse à certains type d'appels (par exemple de type conseil). D'autres possibilités de contrôle sont aussi concevables, comme par exemple, rediriger un appel vers un centre d'appel d'une zone limitrophe.

L'objectif de ce projet de 3$^{\text{ème}}$ année est de formaliser ce problème d'optimisation, de l'étudier aussi bien par des moyens théoriques que par la simulation ou par la mise en œuvre d'algorithmes numériques, afin de pouvoir évaluer et comparer différentes stratégies de conduite, et si possible, d'optimiser celles-ci. On pourra s'appuyer sur un simulateur déjà réalisé.

Le problème ici posé émane de Regis Reboul, en charge du projet de réforme des centres d'appels à la Préfecture de police de Paris. L'encadrement scientifique sera assuré par Xavier Allamigeon et Stéphane Gaubert (INRIA et CMAP). Un doctorant du corps des IPEF, Vianney Bœuf, qui travaille sur cette thématique, sera associé au suivi. Certaines réunions de suivi de ce projet pourront avoir lieu à la BSPP (porte de Champérêt) ou bien à la Préfecture de police (sur l'Ile de la Cité).

Références

X. Allamigeon, V. Bœuf, and S. Gaubert. Performance evaluation of an emergency call center : tropical polynomial systems applied to timed Petri nets. In

FORMATS'15, volume 9268 of *Lecture Notes in Computer Science*. Springer, 2015.

3.15 Optimisation de la couverture des secours en fonction des données météorologiques

Xavier Allamigeon (Xavier.Allamigeon@inria.fr), INRIA – Ecole polytechnique
Stéphane Gaubert (Stephane.Gaubert@inria.fr), INRIA – Ecole polytechnique
Avec la participation de Vianney Perchet (ENSAE).

Ce sujet s'inscrit dans le cadre d'un projet collaboratif ANR "Democrite", avec notamment pour partenaires le CEA, la Préfecture de police, et la Brigade de Sapeurs Pompiers de Paris. L'un des enjeux de ce projet est d'exploiter les données qui sont disponibles (parmi lesquelles l'historique des interventions) afin de construire des indicateurs de risque. Ceux-ci pourraient par la suite être exploités pour suggérer des raffinements des règles d'engagement des moyens de secours (envoi de véhicules issus de différentes casernes aux différentes interventions). Les interventions ont en effet des degrés divers de sévérité et nécessitent des moyens très variables. Aujourd'hui, les règles tiennent compte de différents critères, parmi lesquels le temps de primo-arrivée sur intervention, ou un critère de nature économique.

On aimerait tenir compte également de considérations de couverture des risques. En effet, en situation où les interventions sont nombreuses, il peut se révêler opportun d'aller chercher des moyens plus éloignés pour des interventions non urgentes, afin d'équilibrer la couverture et ainsi de répondre plus rapidement aux interventions les plus urgentes. La météo est l'un des principaux éléments qui influence le nombre d'interventions, et donc le risque.

Dans le cadre de ce projet de 3$^{\text{ème}}$ année, on se propose de construire un modèle d'optimisation stochastique tenant compte de l'aléa météorologique, et de voir, notamment par des expérimentations numériques, dans quelle mesure un algorithme d'allocation des moyens qui tiendrait compte de cet aléa permettrait d'améliorer le service rendu. Ce projet fera appel à des méthodes d'optimisation stochastique ou d'optimisation en ligne, en s'appuyant sur un travail de statistique plus élémentaire. Les expérimentations menées pourront s'appuyer sur des données réelles d'interventions. Vianney Bœuf, qui travaille sur cette thématique, sera associé au suivi. Ce travail pourra donner lieu à une restitution à la BSPP et/ou au sein du projet Democrite.

4 Sciences de la vie

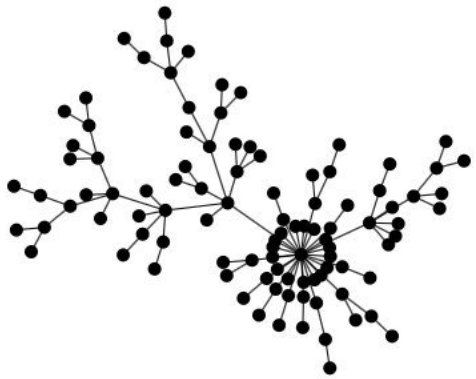
Enseignant référent : Lucas Gerin gerin@cmap.polytechnique.fr

4.1 Interactions entre individus : modèles de renforcement

Mots-clés : Modélisation, Graphes aléatoires, Renforcement, Réseaux sociaux, Propagation d'épidémies

Yule [1] a proposé en 1925 pour la modélisation de l'évolution des espèces un modèle dynamique de croissance avec "renforcement".

Bien plus récemment [2] et dans des contextes différents (modélisation des réseaux sociaux, systèmes complexes), des modèles plus simples de *graphes à attachement préférentiel* ont été introduits. Les sommets de ces graphes sont des individus, et une arête signifie une interaction (amitié, proximité,...) entre deux de ces individus. Ces graphes illustrent également la notion de renforcement : plus un individus a des connections, plus il en aura dans le futur.



Il s'agit d'un domaine très riche au niveau théorique, et aux multiples applications. Selon le goût des élèves, ce projet pourra s'articuler autour des objectifs suivants :

- Étude théorique et numérique du modèle de l'attachement préférentiel. Comparaison avec d'autres modèles.
- Lien avec la modélisation biologique dans l'approche de Yule.
- Modélisation dans les réseaux sociaux, comparaison avec des données réelles.
- Étude de la propagation des épidémies dans de tels modèles.
- Étude d'algorithmes sur des graphes de renforcement.

Une part importante sera consacrée à la simulation, le choix du langage est laissé aux élèves (`python`, `scilab`,...)

Une très bonne référence accessible pour des élèves ayant suivi MAP432 en 2ème année est le livre [3] (Chapitre 8).

Sujet proposé par Lucas Gerin (Ecole Polytechnique, CMAP)
gerin@cmap.polytechnique.fr

Références

[1] G.U.Yule. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis. *Philosophical Transactions of the Royal Society of London.* Series B (1925) 21-87.
[2] A.-L. Barabasi, R.Albert. Emergence of scaling in random networks, *Science* (1999).
[3] R.van der Hofstad. *Random Graphs and Complex Networks* (2015), disponible à http://www.win.tue.nl/~rhofstad.

4.2 Populations en sélection récurrente : effets de la consanguinité et de la liaison génétique sur l'adaptation

Certaines caractéristiques phénotypiques (physiques) des individus d'une espèce donnée dépendent d'un très grand nombre de gènes dont les expressions se combinent. C'est sur les traits phénotypiques que peuvent s'exercer des pressions de sélection (naturelle ou artificielle) favorisant la reproduction de certains individus par rapport à d'autres. C'est notamment ce qu'il se passe lorsqu'on essaie d'optimiser la productivité dans un élevage animalier (par exemple en faisant se reproduire préférentiellement les vaches produisant le plus de lait), ou dans les programmes de production de nouvelles variétés végétales (par exemple en sélectionnant des plantes plus résistantes aux températures élevées et aux faibles précipitations). Grâce au brassage génétique induit par la reproduction sexuée, de telles améliorations peuvent s'accumuler tout au long des générations. Malheureusement, sur le long terme, toute population finie devient consanguine et perd en valeur génétique.

Une toute nouvelle biotechnologie permet d'intensifier le brassage génétique en augmentant le nombre de crossovers méiotiques à chaque génération. Peut-on utiliser cette nouvelle approche pour accélérer la sélection artificielle tout en retardant l'apparition de la consanguinité ? Modéliser l'influence combinée de nombreux gènes lorsque la recombinaison agit et brise les liaisons entre les allèles portés par un individu est complexe et peu d'études s'y sont risquées. Ce projet consistera à explorer cette dynamique complexe via une étude théorique complétée par des approches numériques.

Sujet proposé par Olivier Martin (GQE-Le Moulon) et Amandine Véber (Chercheur CNRS et CMAP).
amandine.veber@cmap.polytechnique.fr

5 Mathématiques Financières

Enseignants référents : Stefano De Marco demarco@cmap.polytechnique.fr, Nizar Touzi touzi@cmap.polytechnique.fr

Il est recommandé de suivre le cours MAP552 Modèles stochastiques en finance pour travailler sur les projets proposés dans cette thématique.

Sujets proposés par :

René Aïd - *EDF*
e-mail : rene.aid@gmail.com
Emmanuel Bacry - *CMAP–Ecole Polytechnique*
e-mail : emmanuel.bacry@polytechnique.fr
Stefano Bosi - *Université d'Evry*
e-mail : sbosi@univ-evry.fr
Yacine Chitour - *Université Paris-Sud & CentraleSupélec*
e-mail : yacine.chitour@lss.supelec.fr
Stefano De Marco - *CMAP–Ecole Polytechnique*
e-mail : demarco@cmap.polytechnique.fr
Laurent Denis - *Université du Mans & CMAP*
e-mail : laurent.denis@univ-lemans.fr
Pierre Henry-Labordere - *Société Générale*
e-mail : pierre.henry-labordere@sgcib.com
Benjamin Jourdain - *CERMICS–Ecole des Ponts ParisTech*
e-mail : jourdain@cermics.enpc.fr
Thibaut Mastrolia - *CMAP–Ecole Polytechnique*
e-mail : thibaut.mastrolia@ceremade.dauphine.fr
Anis Matoussi - *Université du Mans & CMAP*
e-mail : anis.matoussi@univ-lemans.fr
François Pannequin - *ENS Cachan*
e-mail : pannequin@ecogest.ens-cachan.fr
Dylan Possamaï - *Université Paris Dauphine*
e-mail : dylan.possamai@polytechnique.edu
Mathieu Rosenbaum - *LPMA–Université Pierre et Marie Curie*
e-mail : mathieu.rosenbaum@polytechnique.edu

Valorisation d'options, méthodes numériques et de Monte-Carlo

5.1 Produits dérivés sur la volatilité et modèle de variance forward

Les années récentes ont vu l'émergence de produits dérivés d'un nouveau type, dont le pay-off dépend de la volatilité réalisée par un actif de référence sur une période à venir. Ces « swaps de variance » ou « swaps de volatilité » permettent à un investisseur de faire un pari sur le niveau futur de la volatilité sans avoir à émettre de vues sur le niveau des prix eux-mêmes. L'objectif de ce travail sera d'étudier ces instruments et les méthodes proposées pour les évaluer.

Références

[1] P. Carr, D. Madan, *Towards a theory of volatility trading*, in : Volatility, ed. R.A. Jarrow, Risk Publications, 1998.
[2] K. Demeterfi, E. Derman, M. Kamal, J. Zou, *More than you ever wanted to know about volatility swaps*, Journal of Derivatives, Summer 1999.
[3] L. Bergomi, *Smile Dynamics III*, Risk, 90-96, 2008
[3] B. Bühler, *Consistent Variance Curve Models*, Finance and Stochastics, 178-203, 2006.

5.2 Processus de Wishart et modèles à volatilité stochastique multidimensionnels

Les processus de Wishart sont des processus à valeurs dans les matrices symétriques positives. Récemment, ces processus ont été utilisé en finance pour modéliser la covariance instantanée entre différents actifs. L'objectif de ce projet sera dans un premier temps de se familariser avec les processus de Wishart qui ont été indroduits dans [1]. Ensuite, on s'intéressera au modèle proposé dans [3] pour un panier d'actifs. On cherchera à implémenter une méthode de Monte-Carlo permettant de calculer les prix dans ce modèle en s'inspirant de [2] et [4]. On pourra dans un deuxième temps s'intéresser au pricing par méthode de Fourier dans le modèle [3] avec deux actifs.

Références

[1] Bru, M.-F. (1991). Wishart processes. J. Theoret. Probab. 4 725-751.
[2] Ahdida, A. and Alfonsi, A. (2013). Exact and high-order discretization schemes for Wishart processes and their affine extensions. Ann. Appl. Probab., 23(3) :1025-1073.
[3] Da Fonseca J., Grasselli M., and Tebaldi C. (2008). Option pricing when correlations are stochastic : an analytical framework. Review of Derivatives Research.
[4] Ahdida A., Alfonsi A. and Palidda E. (2014). Smile with the Gaussian term structure model. http ://arxiv.org/pdf/1412.7412v2.pdf

5.3 Méthode de Monte-Carlo multipas pour les options européennes

On considère un sous-jacent $(X_t)_{t\in[0,T]}$ dont l'évolution sous la probabilité risque-neutre est donnée par une équation différentielle stochastique (modèle à volatilité locale ou stochastique). En général (en dehors du modèle de Black-Scholes), X_T n'a pas d'expression explicite. Pour calculer le prix $\mathbb{E}\left[e^{-rT}\varphi(X_T)\right]$ d'une option européenne vanille de payoff φ, on approche X_T par X_T^n où $(X^n_{\frac{kT}{n}})_{0\leq k\leq n}$ est un schéma de discrétisation de l'EDS de pas de temps T/n et on calcule la moyenne empirique $\frac{e^{-rT}}{M}\sum_{m=1}^{M}\varphi(X_T^n(m))$ des contributions de M simulations indépendantes de ce schéma. Récemment, Giles [1] a remarqué que combiner plusieurs pas de temps différents dans un estimateur Monte-Carlo multipas permet de réduire le temps de calcul à précision donnée. L'objectif de ce projet est de comprendre l'analyse des erreurs fortes et faibles des schémas de discrétisation d'EDS puis le principe de la méthode de Monte-Carlo multipas et de l'implémenter dans un modèle à volatilité stochastique.

Références

[1] M.B. Giles. Multi-level Monte Carlo path simulation. *Operations Research*, 56(3) :607-617, 2008.
[2] M.B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation, *Annals of Applied Probability*, 24(4) :1585-1620, 2014.
[3] V. Lemaire and G. Pagès. Multilevel Richardson-Romberg extrapolation, preprint arXiv :1401.1177, 2014.

5.4 Estimation de risques VaR dans un modèle incertain

On considère un actif dont le prix satisfait une Equation Différentielle Stochastique dont les coefficients sont mal déterminés. On suppose typiquement que ces différents coefficients prennent leurs valeurs dans des intervalles fixes donnés. On considérera d'abord le cas où l'équation du prix est dirigée par un mouvement Brownien puis le cas où ce prix comporte des sauts modélisés par un processus de Poisson.
On établira des estimations de différents risques VaR et on en déduira également une application au calcul de probabilités de ruine. Pour chaque estimation obtenue, on mettra en oeuvre des simulations numériques afin de tester la précision des estimations obtenues.

Références

[1] Denis L., Fernandez B., Meda A., *Estimation of dynamic var and mean loss associated to diffusion processes*, in Markov Processes and Related Topics : A Festschriftfor Thomas G. Kurtz. J. Feng, S. Ethier, and D. Stockbridge, eds. IMS Collections (4), 301-314 (2008).

[2] Denis L., Fernandez B., Meda A., *Estimation of value at risk and ruin probability for diffusion processes with jumps*, Mathematical Finance Vol. 19(2), pp 281-302 (2009).
[3] Revuz D., Yor M., *Continuous Martingales and Brownian Motion*, springer (1999).

5.5 Modèle à volatilité incertaine et méthode primale pour les BSDEs

Dans le cas où l'on suppose que la volatilité n'est pas connue, l'EDP linéaire de Black-Scholes est remplacée par une EDP non-linéaire. La résolution de cette EDP en grande dimension nécessite des méthodes probabilistes comme les équations stochastiques rétrogrades (BSDEs).
Les objectifs de ce projet sont :

- Comprendre la représentation probabiliste des equations Hamilton-Jacobi-Bellman par des BSDEs.
- Implémenter un schéma numérique *primale* pour un modèle à volatilité incertaine.

Références

[1] J. Guyon, P. Henry-Labordère : *Uncertain volatility model : A Monte-Carlo approach*, Journal of computational Finance, 2011.
[2] A. Fahim, N. Touzi, X. Warin : *A probabilistic numerical method for fully nonlinear parabolic PDEs*, Ann. Appl. Probab. 21, 1322–1364, 2011.
[3] E. Gobet, Lemor, X. Warin : *A regression-based Monte Carlo method to solve backward stochastic differential equations*, Ann. Appl. Probab. 15,3(2005), 2172–2202.

5.6 Options américaines et méthode duale

Une option américaine peut être exercée à tout instant entre la date de son acquisition et son échéance. Pour comprendre si l'option doit être exercée à la date t, il faut comparer le montant récupéré à l'exercice avec la valeur de l'option conditionnelle au fait que l'acheteur décide de ne pas exercer en t. La nécessitée de calculer les espérances conditionnelles rend l'utilisation de la méthode de Monte Carlo particulièrement délicate.
Les objectifs de ce projet sont :

- Comprendre la théorie classique des options américaines et l'algorithme de Longstaff-Schwartz.
- Comparer les méthodes duales.

Références

[1] F. A. Longstaff, E. S. Schwartz, *Valuing American options by simulation : A*

simple least-squares approach, Review of Financial Studies, 14, 113-147, 2001.
[2] Rogers, C. : *Monte-Carlo valuation of American options*, Mathematical finance 12 (3), 271-286.
[3] Broadie, M, Cao, M. : *Improved lower and upper bound algorithms for pricing American options by simulation.*
[4] P. Glasserman, *Monte Carlo methods in financial engineering*, Chap 8, Springer.

Liquidation optimale avec impact sur les prix

5.7 Stratégies optimales de passage d'ordre et estimation online de l'impact

Lorsque des ordres sont passés en nombre élevé ou pour des volumes importants en peu de temps, ils génèrent un impact de marché : le prix du sous-jacent est impacté à la hausse s'il s'agit d'achats, à la baisse s'il s'agit de ventes, et l'ordre global est en généralement exécuté à un prix différent du meilleur prix observé. Il faut évidemment en tenir compte.

Malheureusement, il est très délicat de mesurer cet impact : il peut évoluer selon les conditions de marché, et surtout on ne peut l'observer qu'en passant des ordres, ce qui peut générer des pertes importantes si le paramètre d'impact est mal calibré a priori. On se trouve donc dans une situation classique : on veut agir de manière optimale sur un système, ceci nécessite la connaissance de paramètres, mais on ne peut les estimer quand agissant sur le système.

Une solution consiste à se donner un prior sur la distribution du paramètre d'intérêt et à réviser cette loi a priori de manière bayésienne à chaque fois que l'on observe l'impact réalisé.

L'objectif de ce projet est de mettre en place cette approche dans différents cadres d'application. On s'intéressera bien évidemment aux modèles avec impact sur les prix, mais également aux stratégies de placement d'ordre dans un carnet, ou aux choix optimal d'un dark pool.

Références

[1] N. Baradel, B. Bouchard, and M. Dang. Optimal control of trading algorithms and bayesian parameters adjustments. preprint.
[2] B. Bouchard, M. Dang, and C.-A. Lehalle. Optimal control of trading algorithms : A general impulse control approach. *SIAM Journal on financial mathematics*, 4 :404–438, 2011.
[3] J.-P. Bouchaud. Price impact. *Encyclopedia of quantitative finance*, 2010.
[4] D. Easley and N. M. Kiefer. Controlling a stochastic process with unknown parameters. *Econometrica : Journal of the Econometric Society*, pages 1045–1064, 1988.
[5] O. Guéant, C.-A. Lehalle, and J. Fernandez-Tapia. Optimal portfolio liquidation with limit orders. *SIAM Journal on Financial Mathematics*, 3(1) : 740–764, 2012.

[6] S. Laruelle, C.-A. Lehalle, and G. Pagès. Optimal posting price of limit orders : learning by trading. *Mathematics and Financial Economics*, 7(3) :359–403, 2013.
[7] C.-A. Lehalle and S. Laruelle. *Market microstructure in practice*. World Scientific Publishing Co. Pte. Ltd., 2013.

Problèmes Principal-Agent

5.8 Introduction à la théorie des contrats

Prenons l'exemple de deux parties coopérant, l'une appelée *le Principal* propose à la seconde appelée *l'Agent* de gérer un projet risqué. En échange de l'effort qu'il fournit dans son travail, l'Agent reçoit de la part du Principal une partie de la valeur du projet en guise de salaire. Dans cette situation, l'effort de l'Agent a des conséquences directes sur la valeur du projet. Plusieurs situations, dépendant de l'information disponible pour le Principal, sont alors envisagées. Tout d'abord, lors de la proposition du contrat, le Principal assure à l'Agent une part du projet et lui impose un niveau d'effort à fournir, ce qui est nommé dans la littérature le partage du risque. Imaginons maintenant une seconde situation où le Principal propose à l'Agent une part du projet mais ne contrôle pas l'effort fourni par ce dernier. Ce genre de situation est plus communément appelé l'aléa moral et fut initialement étudié dans [**?**]. Dans chacune des situations énoncées, le but du Principal est de proposer un contrat à l'Agent lui permettant de maximiser son utilité comme fonction de la valeur terminale du projet.

Le but de ce travail sera de résoudre un problème simple sous aléa moral, s'apparentant à un équilibre de Stackelberg, en deux temps :

1. On commence par résoudre, à contrat fixé, le problème de l'Agent. On obtient ainsi son effort optimal étant donné un contrat proposé par le Principal.
2. On injecte dans le problème du Principal l'effort de meilleure réponse de l'Agent précédemment trouvé, et on résout le problème du Principal, en fournissant le contrat optimal proposé à l'Agent.

On supposera ici que la valeur du projet subit un effet d'entrainement de paramètre α dont la dynamique est donnée par

$$dX_t = (a_t + \alpha X_t)dt + \sigma dW_t^a,\ X_0 \in \mathbb{R}.$$

Le projet sera illustré par des simulations numériques du contrat optimal proposé par le Principal et de l'effort optimal de l'Agent suivant différentes valeurs de α.

Les différents outils abordés seront des résultats fondamentaux issus

- du calcul stochastique : formule d'Itô, théorème de représentation des martingales, formule de Feynman-Kac,
- de la théorie du contrôle : équations d'Hamilton-Jacobi-Bellman.

Références

[1] Cvitanić, J., Zhang, J. (2012). Contract theory in continuous time models,

Springer-Verlag.
[2] Holmström B., Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives, *Econometrica*, 55(2) :303–328.

Calibration de modèle

5.9 Calibration d'un triangle de taux de change

Le modèle simple de Garman-Kohlhagen permet de comprendre le pricing des options portant sur un taux de change. Il est alors facile de généraliser la formule de Dupire [1] pour construire un modèle à volatilité locale qui reproduise les prix de marché des options liquides portant sur ce taux.
Lorsque l'on considère maintenant deux taux de change impliquant deux fois la même devise, par exemple DOL/EUR et GBP/EUR, le problème de la calibration devient beaucoup plus complexe : en effet le quotient du premier taux par le second est lui-même le taux de change GBP/DOL. Il s'agit alors de construire un modèle compatible avec les prix d'options liquides portant sur les deux taux de départ DOL/EUR et GBP/EUR mais aussi sur leur rapport GBP/DOL.
Le but de cet EA est de comprendre la calibration d'un seul taux de change avant d'étudier l'approche proposée dans [2] pour calibrer deux taux de change.

Références
[1] B. Dupire, *Pricing with a smile*, Risk, January 1994.
[2] J. Guyon, *A new class of local correlation models*, preprint SSRN, 2013.

5.10 Calibration d'un modèle hybrique taux-action par une méthode particulaire

La calibration de modèles diffusifs sur des smiles de marché se formalise comme la résolution numérique d'une équation de Fokker-Planck non-linéaire. La résolution de cette EDP en grande dimension nécessite des méthodes probabilistes comme les méthodes particulaires pour les EDS non-linéaires de MCKean.
Les objectifs de ce projet sont :

- Comprendre la théorie classique des EDS non-linéaires de McKean et l'algorithme particulaire.
- Implémenter un schéma numérique pour la calibration d'un modèle hybrique taux-action.

Références
[1] J. Guyon, P. Henry-Labordère : *The Smile Calibration Problem Solved*, Risk magazine (2012).

Marchés d'énérgie

5.11 Modélisation structurelle des prix de l'électricité et interaction stratégique.

Une hypothèse importante de la théorie de l'évaluation des actifs contingents est la stockabilité du sous-jacent. L'électricité n'étant pas stockable, l'évaluation de ses dérivés nécessite d'adapter ce cadre. Ces dernières années, des modèles de prix de l'électricité fondés sur une représentation simplifiée de l'équilibre offre-demande ont permis à la fois une définition précise de l'absence d'opportunité d'arbitrage et une évaluation efficace des dérivés les plus simples (futures, options spreads) [1]. De nombreuses variantes ont été proposées dans la littérature [4,5]. L'objet de ce projet est d'étudier la possibilité de tenir compte des effets stratégiques sur le marché au comptant et/ou à terme entre différents agents [2,3].
Les étapes requises sont :

- l'étude des modèles de prix structurels de l'électricité
- la formulation d'un cadre d'intéraction stratégique entre acteurs permettant un calcul analytique des futures
- l'étude et la simulation du modèle

Références

[1] R. Aïd, L. Campi, N. Langrené. A structural risk-neutral model for pricing and hedging power derivatives. *Mathematical Finance*. 23(3) :387–438. 2013.
[2] B. Allaz. Oligopoly, uncertainty and strategic forward transactions. *Internat. J. Indust. Organ.*, 10(2) :297–308, 1992.
[3] B. Allaz, J.-Y. Villa. Cournot Competition, Forward Markets and Efficiency. *Journal of Economic Theory*, 59(1) :1–16, 1993.
[4] R. Carmona and M. Coulon. A Survey of Commodity Markets and Structural Models for Electricity Prices. *Quantitative Energy Finance : Modeling, pricing and hedging in energy and commodity markets*, p. 41–83. F. E. Benth, V. Kholodny and P. Laurence, eds. Springer edition, 2013.
[5] R. Carmona, M. Coulon, and D. Schwarz. Electricity price modeling and asset valuation : a multi-fuel structural approach. *Mathematics and Financial Economics*, 7(2) :167–202, 2013.

5.12 Quand faut-il construire une centrale électrique ?

La règle la plus utilisée par les entreprises du secteur électrique pour décider de leurs investissements en moyens de production est la positivité de la valeur actuelle nette. Toutefois, depuis les années 80, la théorie des options réelles propose d'abandonner ce principe au profit d'une règle représentant les investissements comme des options américaines : l'entrepreneur peut choisir le moment qui maximise la valeur actuelle nette au lieu de se limiter à sa positivité [4].
Cette approche conduit à des équations aux dérivées partielles à frontières libres. Dans certains cas simples, des solutions analytiques peuvent être obtenues. Mais,

la prise en compte de caractéristiques de la centrale comme le délai de construction ou de propriétés du prix de marché comme le retour à la moyenne rendent rapidement nécessaire l'utilisation de méthodes numériques.
Le but de ce projet est de déduire quelques conséquences concrètes de la règle de décision d'investissement fondée sur la théorie des options réelles au cas des investissements en production d'électricité.
Les étapes requises sont :

- l'etude des bases de la théorie des options réelles [1,3]
- la formulation d'un cadre de modélisation de l'actif (la centrale de production) et du marché (modèle de prix) présentant un compromis entre difficulté de résolution et réalisme,
- la résolution de leur modèle ou bien de façon analytique ou bien de façon numérique [2]
- la formulation de règles d'investissement en fonction du prix de marché et des caractéristiques de la centrale,
- la formulation d'une analyse critique de leur modèle.

Références

[1] R. Aïd. *A review of optimal investment rules in electricity generation.* Qantitative Energy Finance, F. E. Benth, P. Laurence, V. Kolodnyi eds, Springer, 2013.
[2] R. Carmona, P. Del Moral, N. Oudjane et P. Hu. *Numerical Methods in Finance.* Springer, 2012.
[3] A. Dixit and R. Pindyck. *Investment Under Uncertainty.* Princeton University Press, 1994.
[4] R. McDonald and D. Siegel. The value of waiting to invest. *Quarterly Journal of Economics*, 101 :707–727, 1986.

Délit d'initié

5.13 Délit d'initié : modélisation et détection

En France, l'AMF (Autorité des Marchés Financiers) est chargée de surveiller les opérations boursières. Depuis une dizaine d'années, ces autorités de surveillance ont fait beaucoup de progrès dans la détection de comportement initié grâce notamment à de meilleures techniques de surveillance. Un exemple de grande envergure a eu lieu entre novembre 2005 et mars 2006, période durant laquelle 10 millions de titres EADS ont éte vendus, pour une plus value de près de 90 millions d'euros. Ces mouvements anormaux ont été décelés par l'AMF, et ont conduit à des enquêtes sur 21 hauts dirigeants d'EADS.
L'objectif des deux sujets suivants est d'étudier une modélisation d'un délit d'initié et de mettre en place un test de détection. On se placera dans le cadre d'un marché financier dont les prix des actifs sont dirigés par un mouvement brownien. L'information minimale dont disposent les agents pour résoudre leur problème d'optimisation est celle obtenue par l'observation du processus des prix. Cepen-

dant, il semble que les agents sont informés de manière hétérogène et reçoivent un flux d'information qui leur est propre. Pour de tels initiés, on étudiera :
- Les problèmes d'arbitrage et de réplication d'actifs risqués,
- Le gain de l'initié (par rapport à un non-initié),
- La mise en oeuvre de tests de détection.
On peut considérer plusieurs modélisations de l'information privée. Nous étudierons le cas où l'investisseur possède une information initiale, i.e. il connait, dès l'instant $t = 0$, une fonctionnelle des trajectoires du processus des prix. La clé de cette modélisation est la théorie du grossissement initial de filtration par une variable aléatoire. On étudiera par exemple le cas où l'initié connait le ratio du prix terminal de deux actifs, ou bien encore le cas où l'initié sait si le prix terminal d'un actif sera dans une fourchette donnée ou non.

Références
[1] A. Grorud, M. Pontier, *Comment détecter le délit d'initié*, C.R. Acad. Sci. Paris, t. 324, p. 1137-1142, 1997.
[2] A. Grorud, M. Pontier, *Insider trading in a continuous time market model*, International Journal of Theoretical and Applied Finance, 1, p. 331-347, 1998.
[3] M. Pontier, *Modélisation et détection du délit d'initié*, Matapli 77, 2005.

Ambiguity in Finance and Insurance

5.14 Ambiguity and Macro-Finance

Keywords : General equilibrium theory, ambiguity theory, rational bubbles.

Highlights : The existence of dynamic general equilibrium when agents have heterogeneous beliefs and financial frictions take place. How to evaluate the equilibrium price of an asset and how to determine its fundamental value ? While Harrison and Kreps (1978), Werner (2014) provide some examples of speculative bubbles, an analysis of dynamics of bubbles remains. What are the effects of ambiguity and beliefs heterogeneity of agents on the aggregate level of investment and the rate of growth in the real economy ? Is there any relation between the existence of speculative bubbles and the efficiency of productive sector in presence of ambiguity. How is about the interplay between the dynamics of bubbles and the economic growth ? Under which conditions bubbles help/harm growth ?

Issue : Theorists are paying a growing attention to the general equilibrium implications of agents heterogeneity. Different aspects of heterogeneity are usually considered : endowments, technologies, preferences. Instead, this doctoral project aims at focusing on the role of heterogeneity in beliefs and its consequences on asset prices and economic growth when financial market imperfections are at work. More precisely, we are interested in the occurrence of speculative bubbles, their dynamic properties and effects on capital accumulation. Their interference with capital accumulation may be beneficial or detrimental depending whether agents oversave or not.

The emergence of asset price bubbles and their sudden bursting may have a dramatic impact on economic activities. In the last decades, theorists faced many (mathematical) obstacles to provide a realistic picture of the interplay between financial markets and the real economy. To understand the real impact of financial crisis and speculative attacks, a representation of markets interactions is suitable, that is a general equilibrium perspective. Since the early Eighties, theorists succeed in modelling the (so-called rational) bubbles resulting from the optimal choice of individuals, and introducing them in general equilibrium models.
Today, two important challenging issues are addressed by them : the role of agents heterogeneity in bubble formation and the growth effects of bubbles. There is a huge literature on rational bubbles in general equilibrium models without ambiguity. Tirole, Kocherlakota, Santos and Woodford study rational financial asset bubbles while Bosi, Le Van and Phamfocus on rational physical capital, land, and housing bubbles. It is known that rational bubbles fail in economies populated by identical agents. Agents' heterogeneity is a necessary but not a sufficient condition for rational bubbles to exist. In Bosi, Le Van and Pham, the development level of the financial system is characterized by the maximum ratio of the investors' debt to their collateral values. They have shown that a financial system "good enough" rules out any bubble, even in presence of many heterogeneous investors. Otherwise, there is room for bubbles.
While the existing literature has given us a number of insights for why rational bubbles appear, we know much less about the mechanism of speculative bubbles. In this project, we are especially interested in the general equilibrium effects of ambiguity (beliefs heterogeneity) and the interplay between financial imperfections and capital accumulation.
The notion of ambiguity is unambiguously and formally given as follows. Let each agent i know the set $\mathcal{P}$ of probabilities about fundamentals, for example, the dividend process, or the endowments she receives. Agent i will solve the following problem by choosing the pair (x, P).

$$\max_{x \in X_i} \left[\min_{P \in \mathcal{P}} u_i \left(x, P \right) \right]$$

where x is the allocation of agent i, X_i denotes her set of feasible allocations and u_i her utility.
In a first stage, we will assume the same $\mathcal{P}$ for any agents. In a second stage, we will consider the case of different $\mathcal{P}$: each agent i has her own set $\mathcal{P}$ of probabilities. Of course, the second case is much more difficult to deal with but it would provide richer implications.
In this first part of the research project, we would like to study the following questions :

1. The existence of dynamic general equilibrium when agents have heterogeneous beliefs and financial frictions take place.

2. How to evaluate the equilibrium price of an asset and how to determine its fundamental value ?

3. The impact of ambiguity on the existence of and the size of bubbles.
The second part of this research project will focus on the impact of ambiguity and bubbles on economic growth. Let us sketch the basic idea. If investors expect more uncertainty in the future, they may reduce their investments and promote a slowdown of economic activities (firms face a difficulty to fund their production projects). Many promising questions will be tackled :

1. What are the effects of ambiguity and beliefs heterogeneity of agents on the aggregate level of investment and the rate of growth in the real economy ?

2. Is there any relation between the existence of speculative bubbles and the efficiency of productive sector in presence of ambiguity.

3. How is about the interplay between the dynamics of bubbles and the economic growth ? Under which conditions bubbles help/harm growth ?

Références

[1] Bosi S., C. Le Van and N-S. Pham, Intertemporal equilibrium with production : bubble and efficiency, EPEE working paper (2014).
[2] Bosi S., C. Le Van and N-S. Pham, Intertemporal equilibrium with heterogeneous agents, endogenous dividends and borrowing constraints, EPEE working paper (2015).
[3] Harrison J. M., Kreps M. D. Speculative investor in a stock market in heterogeneous expectations, The Quarterly Journal of Economics 92, 323-336 (1978).
[4] Kocherlakota, N. R., Bubbles and constraints on debt accumulation, Journal of Economic Theory, 57, 245 - 256 (1992).
[5] Le Van, C., Pham, N.S., Vailakis, Y., Financial asset bubble with heterogeneous agents and endogenous borrowing constraints, Working paper, 2014.
[6] Le Van, C., and Pham, N.S., Intertemporal equilibrium with financial asset and physical capital, Economic Theory (forthcoming 2015).
[7] Santos, M. S., and Woodford, M. Rational asset pricing bubbles, Econometrica, 65, 19-57 (1997).
[8] Tirole J., Asset bubbles and overlapping generations, Econometrica 53, 1499-1528 (1985).
[9] Werner J., Speculative trade under ambiguity, mimeo (2014).

5.15 Ambiguity and Insurance

Keywords : Insurance, ambiguity theory, experimental economics.

Purpose : This project aims at representing the behavior of risk-lover individuals facing different insurance contracts beyond the conventional view (only risk-averse individual are interested in insurance) and at providing an empirical validation of this theory through an experimental approach.

Issue : This project aims at a better understanding of insurance behavior by taking into account the heterogeneity of individuals regarding risk and ambiguity attitudes. It requires a strong interaction between mathematical modeling, behavioral economics, and experimental economics.
Indeed, insurance theorists are polarized only on the assumptions of risk aversion

or ambiguity aversion, due to the against-intuitive nature of the alternative hypothesis. However, since Kahneman and Tversky , behavioral observations have highlighted the likelihood of risk loving. Several recent contributions argue for an extension of economic analysis to the case of risk loving (Crainich, Eeckhoudt and Trannoy (2013), Corcos, Pannequin and Montmarquette). It seems obvious that a risk lover does not want to buy any insurance coverage. However, studying the behavior of risk lovers is relevant because most of the insurance contracts are mandatory in real life.

In his seminal article, Raviv (1979) proved and extended the well-known result of Arrow (1965) – the optimality of a deductible – but those results were based on the assumption of risk aversion. So, a first extension of the canonical model of Raviv (1979) will be devoted to the characterization of optimal contracting for risk lovers.

A second way of extending the model Raviv (1979) will rely on the assumption of ambiguity. Using the decision model of "smooth ambiguity" of Klibanoff et al. (2005, 2009), Gollier (2012) states that in the presence of ambiguity aversion, the optimal contractual form is conditioned by the structure of ambiguity. Contrary to intuition, ambiguity aversion does not necessarily mean an increase in the demand for insurance. Again, it seems relevant to extend these results to account for the heterogeneity in ambiguity attitudes.

The experimental study of Roy and Chakravarty shows that risk-averse subjects are not necessarily ambiguity averse. It is entirely plausible that a high proportion of individuals is characterized by both risk aversion and ambiguity-loving or by both risk loving and ambiguity aversion.

These facts motivate to extend the model Raviv to the different combinations of attitudes – aversion/loving with risk/ambiguity - as theorists have addressed only one case out of four (risk aversion and ambiguity aversion). From a theoretical point of view, this generalization of the model is a challenge. This will require handling sophisticated mathematical tools (developments in geometric control and nonconvex analysis).

The theoretical predictions will be experimentally tested in the laboratory. Several experimental protocols will be developed to test the behavioral predictions for insurance demand under risk and ambiguity. In each case, risk and ambiguity attitudes of the subjects will be elicited in a first step (according to the methods of Holt and Laury and Chakravarty and Roy). In a second step, insurance choices will be analyzed in the Lab, both for risk and ambiguity settings.

Références

[1] Chakravarty, S. and J. Roy (2009) "Recursive expected utility and the separation of attitudes towards risk and ambiguity : an experimental study", Theory and Decision, 66(3), 199-228.

[2] Corcos, A., Pannequin, F. and C. Montmarquette (2013a), “Risk lovers also love insurance”, WP CIRANO, Montréal.

[3] A. Corcos, C. Montmarquette and F. Pannequin (2013b), Assurance et autoassurance : une étude expérimentale, Rapport pour le Ministère de l’Enseignement

supérieur, de la Recherche, de la Science et de la Technologie du Québec, CIRANO, Montréal.
[4] Crainich, D., L. Eeckhoudt & A. Trannoy (2013), "Even (mixed) Risk Lovers Are Prudent", American Economic Review.
[5] Gollier, Ch. (2012), “Optimal insurance design of ambiguous risks”, TSE, LERNA.
[6] Holt, C. and S. Laury (2002). Risk Aversion and Incentive Effects, American Economic Review 92 (5), 1644- 1655.
[7] Klibanoff, P., M. Marinacci and S.Mukerji (2005) “A smooth model of decision making under ambiguity”, Econometrica,73(6), 1849-1892.
[8] Klibanoff, P., M. Marinacci and S.Mukerji (2009) “Recursive smooth ambiguity preferences”, Journal of Economic Theory,144, 930-976.
[9] Pannequin, F., Corcos, A. and C. Montmarquette (2013), “The global accounting heuristic of insurance and self-insurance demands”, WP CIRANO, Montréal.

Stratégies haute fréquence, microstructure des marchés

La disponibilité de données haute fréquence, la multiplication des places de marchés, ainsi qu'une compréhension de plus en plus fine des phénomènes de microstructure, ont ouvert de nouvelles perspectives en finance de marché. En particulier, le trading haute fréquence est né de la volonté d'optimiser les transactions en profitant de ce nouveau contexte. Son essor récent a nécessité le développement de méthodes originales de mathématiques financières et de statistique des processus. Un nombre grandissant d'équipes de trading propriétaires, de salles de marchés et de hedge funds y ont aujourd'hui constamment recours.

5.16 Estimation du Market impact à partir de données haute fréquence

L'estimation de la courbe de market impact d'un meta-ordre (ordre composé de plusieurs petits ordres) est essentielle dans la problématique d'exécution optimale. Il s'agit de quantifier, la variation moyenne du prix à partir du début de l'exécution de ce meta-ordre. Cette courbe présente généralement deux phases : une phase ascendante où le prix monte (dans le cas d'un ordre d'achat) et, dès que l'exécution du meta-ordre est fini, une phase de relaxation. L'estimation, généralement obtenue à l'aide de données clients (pour identifier le meta-ordre), est excessivement difficile et bruitée. Des modèles (processus ponctuels) de prix et flux d'ordres haute fréquence sur des données de marché (anonymes) peuvent être utilisées pour stabiliser grandement cette estimation. Dans le cadre de ce projet, les modèles à base de processus ponctuels auto-excitants seront étudiés.

Références

[1] *Point spectra of some mutually exciting point processes*, A.G. Hawkes, Biometrika, 58 (1971).
[2] *The Non-Linear Market Impact of Large Trades : Evidence from Buy-Side Order Flow*, Bershova, N. and Rakhlin, D., Social Science Research Network Working Paper Series, 2013
[3] *Hawkes model for price and trades high-frequency dynamics*, E. Bacry, J.F. Muzy, Quantitative Finance.

5.17 Estimation de la volatilité historique dans un cadre multifractal

Les premiers modèles multifractals pour la volatilité historique ont été introduits dès la fin des années 60 par Mandelbrot. Depuis, ils ont fait l'objet de très nombreuses recherches académiques. Ainsi, par exemple les modèles MRM (Multifractal Random Measure) [1]permettent de reproduire de façon précise de nombreux faits empiriques établis sur une grande gamme déchelle de temps (de l'heure à l'année) : non gaussianité (resp. gaussianité) à petite (resp. grande) échelle de temps, effet de clustering, scaling des moments en loi de puissance, ... Ces modèles permettent notamment de réaliser des prédicteurs de volatilité historique multiéchelles particulièrement performants.
Récemment un modèle de "rough volatility" a été introduit [2]. Ce modèle n'est pas multifractal à proprement parler, mais il a des propriétés extrêmement proches. Il s'agira dans ce projet de comparer les deux approches et de comprendre dans quelle mesure l'un ou l'autre modèle est plus performant pour l'estimation ou pour la prédiction.

Références

[1] *Log-Normal continuous cascades : aggregation properties and estimation. Application to financial time-series*, E.Bacry, A.Kozhemyak, J.F.Muzy, Quantitative finance, Volume 13, Issue 5, pp 795-818 (2013)
[2] *Volatility is rough*, T.Jaisson, M.Rosenbaum, arXiv :1410.3394

5.18 Détection de choc sur données financières à haute fréquence

La dynamique des marchés financiers se caractérise par un jeu subtil entre les effets endogènes et exogènes. La dynamique des prix des actifs, en fait, est affectée par l'arrivée de « news » (« nouvelles ») exogènes, qui modifient l'évaluation du juste prix de l'actif, et par les actions des agents du marché eux-mêmes. Parfois, l'arrivée d'un nouvel élément d'information peut avoir un très grand impact sur le prix et sur la volatilité d'un actif. La détection et la caractérisation de ces éléments de surprise en séries financières [1-3] est d'une grande importance pour la gestion des risques, ainsi que pour les décisions et les algorithmes de trading [4] et de

nombreux travaux de recherche portent sur ces sujets.
Dans le cadre de ce projet, on devra d'abord comprendre les outils de détection des « chocs » et ensuite les appliquer à des données financières haute fréquence. On peut alors tenter de répondre à plusieurs questions de recherche telles que

- Combien de chocs sont détectés par différents outils et combien d'entre eux peuvent être attribués à de « vraies news » (e.g., annonce du chiffre du chômage, etc...) par opposition à des mécanismes purement endogènes ?
- Les sauts détectés sont-ils similaires pour les différents actifs ?
- Y a-t-il des évidences d'anticipation par les agents de ces news, en particulier lorsque l'heure du communiqué de la news est prévue ?
- ...

Références

[1] Lee, S. S. and Mykland, P. A. [2] Joulin, A., Lefevre, A., Grunberg, D., and Bouchaud, J. P. *Stock price jumps : News and volume play a minor role.* Wilmott Magazine, 1-7 (2008).
[3] Rambaldi, M., Pennesi, P., and Lillo and F. *Modeling foreign exchange market activity around macroeconomic news : Hawkes-process approach.* Physical Review E 91.1 (2015) : 012819.
[4] Groß-Klußmann, A., and Hautsch, N. *When machines read the news : Using automated text analytics to quantify high frequency news-implied market reactions.* Journal of Empirical Finance, 18(2), 321-340 (2011).

5.19 Estimation haute fréquence de la volatilité, application au trading d'options

Dans ce projet, on se placera dans la situation d'un trader haute fréquence souhaitant faire de l'arbitrage sur options. L'idée est de détecter les anomalies de valorisation en comparant prix d'options et mesures de volatilité. On montrera dans un premier temps que le cadre usuel d'une modélisation brownienne est insuffisant dans ce contexte de données haute fréquence. On s'appuiera ensuite sur différentes extensions de ce cadre pour modéliser la microstructure des marchés et construire des stratégies de trading.

Références

[1] Y. Aït-Sahalia, P. A. Mykland, L. Zhang, *A tale of two time scales : Determining integrated volatility with noisy high frequency data*, JASA 77, 100(472), 1394-1411, 2005.
[2] F. G. Bandi, J. R. Russell, C. Yang, *Realized volatility forecasting and option pricing*, Journal of Econometrics, 147(1), 34-46, 2008.
[3] O. E. Barndorff-Nielsen, P. Hansen, A. Lunde, N. Shephard, *Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise*, Econometrica, 2008.

5.20 Corrélation haute fréquence, application au market impact

Bien que la présence de corrélations haute fréquence soit un consensus de marché et que de nombreuses stratégies en place (comme le pair trading) optimisent un critère multi-titres, la mesure et l'exploitation des dépendances entre les variations des prix de deux titres ont été peu explorées. On montrera tout d'abord que, même dans le cadre brownien, la simple asynchronicité des prix (les transactions n'ont pas lieu aux mêmes instants pour deux actifs différents) explique en partie l'effet Epps, c'est à dire une estimation haute fréquence des corrélations systématiquement proche de zéro. On montrera comment corriger cet effet puis on tentera de résoudre le problème dans un cadre plus réaliste, permettant de reproduire les effets de microstructure des marchés. On appliquera les résultats obtenus à l'optimisation de la vente d'un portefeuille d'actifs.

Références

[1] R. Almgren, N. Chriss, *Optimal execution of portfolio transactions*, J. Risk 3, 5-39, 2000.
[2] T. Hayashi, N. Yoshida, *On covariance estimation of non-synchronously observed diffusion processes*, Bernoulli 11(2), 359-379, 2005.
[3] L. Zhang, *Estimating covariation : Epps effect, microstructure noise*, Journal of Econometrics, 160(1), 33-47, 2011.