

Trouver un titre au projet PHC-Utique*

Attention ! Ceci n'est pas forcément le plan de présentation du projet. Je pense que la Section 7 devrait être dissoute dans les Sections 4, 5 et 6. Chacun devra compléter puis valider sa thématique.

Domaine de recherche : Mathématiques et leurs interactions ;

Mots clés : Systèmes polynomiaux, Plans d'expériences, Réseaux de neurones profonds ; Modélisation ; Plans d'expériences ; Big-Data ;

Discipline : Mathématique ; BioInformatique ; Mathématiques Appliquées ; Intelligence Artificielle ;

Contents

1	Résumé succinct du projet	2
2	Principaux objectifs	3
3	Présentation de chaque équipe engagée	5
3.1	Identification des participants en France	5
3.2	Identification des participants en Tunisie	5
3.2.1	Faculté des Sciences de Sfax	5
3.2.2	Laboratoire BIMS de l'Institut Pasteur de Tunis	6
3.3	Coopérations antérieures des partenaires	6
4	Données biologiques et médicales	7
4.1	Systèmes dynamiques discrets	7
4.2	Plan d'expériences	9
4.3	Big-data - Données Massive	11
4.3.1	ce qui est proposé (voir mail)	11

*Systèmes Polynomiaux et Applications en Biologie et Traitement du Signal

4.3.2	Réponses à mes questions (voir mail)	11
5	Données Floues et Plan d'expériences	12
5.1	Données Floues	12
5.2	Traitement des données mixtes	13
6	Planification des résultats attendus	14
6.1	Equipe tunisienne	14
6.1.1	Intérêts scientifiques	14
6.1.2	Retombées socio-économiques	15
6.2	Equipe française	15
6.2.1	Intérêts scientifiques	15
6.2.2	Retombées socio-économiques	15
7	Besoins en matériel/séjours	15

1 Résumé succinct du projet

Il nous faut construire un discours autour des données médicales : plans d'expériences, résolution avec des données floues et incertaines (ça c'est pour Philippe et Annick) : Fuzzy Numbers réseaux de neurones ("deep learning" car maintenant c'est cela qui se dit pour "vendre" ... subitement tout le monde fait du "deep learning" et devient important, dans le vent, etc).

Calcul Formel (résolutions algébriques), ça c'est notre coeur de métier ! alors on le glisse ... systèmes dynamiques discrets, méthodes algébro/numérique ? sécurisation des données personnelles (Pierre Saurel).

Peu sont en capacité de réunir tous nos atouts en se connaissant si bien (plusieurs collaborations existantes antérieures au projet): Inst. pasteur Tunis (médical ; Cancer), Calcul Formel, Fuzzy, Plans d'expériences, systèmes dynamiques discrets et IA.

Ines a mis l'accent sur l'utilisation de SageMath/Python/R avec l'excellent cloud cocalc.com. Mais attention, les ressources sont limitées et il faut penser à acheter des espaces de calculs pour des machines GPU supportant de grosses BD... Pareil concernant les besoins : penser à l'acquisition de cartes GPU pour notre serveur de calcul à l'IPT...

2 Principaux objectifs

Tunisie : - Bases de Gröbner sur les entiers, Bases de Gröbner dynamiques, algorithme dynamique de Buchberger, algorithmes pour la réduction des matrices polynomiales unimodulaires, - Théorie de l'élimination, Théorie des invariants, - Traitement du signal, Réseaux de neurones profonds et Apprentissage - Modélisations mathématiques en biologie, réseaux cellulaires biologiques, biologie algébrique. - Développeurs/Utilisateurs des systèmes de Calcul Formel

France : - Résolutions efficaces de systèmes polynomiaux : symboliques/numériques, variétés polaires, systèmes triangulaires - Algorithmes pour la réduction des matrices polynomiales unimodulaires, Théorème de Quillen-Suslin - Plans d'expérience - Théorie de l'élimination, Théorie des invariants - Théorie de Galois et systèmes polynomiaux galoisiens - Statistique et systèmes biologiques, - Développeurs/Utilisateurs des systèmes de Calcul Formel

La complexité croissante du vivant et la masse de données générées par la mise en évidence d'études issues de différentes branches de la biologie, ont érigé la modélisation en approche privilégiée pour intégrer ces connaissances dans un esprit fortement pluridisciplinaire. De l'individu au gène en passant par la cellule, la modélisation permet de décrire l'ensemble des interactions existantes et d'appréhender les mécanismes à l'origine des phénomènes biologiques observés. C'est autour de ce souci d'intégrer l'ensemble des niveaux d'organisations du vivant que s'articulent nos principaux axes de recherche : de la population à l'individu (Epidémiologie et Evolution) ; de l'individu à la cellule (Modélisation et compréhension de la réponse immune) ; de la cellule au gène (Biosystemique visant à élucider les mécanismes induits par la maladie/infection ou des mécanismes de résistance et de susceptibilité). Ce dernier axe de recherche comprend aussi bien la maîtrise de la collecte des données, que leur intégration ou leur traitement.

Aussi, contrairement à l'analyse classique qui a pour but la compréhension du fonctionnement de quelques composants cellulaires, l'analyse systémique se propose d'étudier la cellule dans sa globalité. Pour cela il est nécessaire d'intégrer les informations relatives à un très grand nombre de composants afin d'en déduire les principales voies métaboliques, signalétiques, énergétiques, reproductives et leur interactions fonctionnelles. Cette discipline très récente permet une vision holistique du fonctionnement cellulaire, plus proche de la complexité de la situation réelle dans les conditions physiologiques ou pathologiques : cancérogenèses, maladies génétiques mono-factorielles

et multifactorielles, maladies infectieuses, etc.

Depuis ces 20 dernières années, les développements très rapide de la génomique fonctionnelle, de la transcriptomique et de la protéomique ont permis une accumulation exceptionnelle de données relatives à la séquence entière de génomes d'organismes, l'expression globale de ces gènes (transcriptome) et à la modification des protéines qui en découle (protéome). La plupart de ces données sont accessibles sur les bases de données internationales et peuvent être valorisées dans la mesure où des outils d'analyses sont disponibles et maîtrisés. Chaque développement nouveau de la biologie implique le développement nouveau d'un outil d'analyse. En effet, les outils d'intelligence artificielle, se sont démocratisés par un accès rapide à des machines de calculs de plus en plus puissantes, avec des capacités de stockages pratiquement illimitées. Cette facilité de calculs a engendré des avancées considérables dans la classification (supervisée ou pas) et les simulations in-silico via des réseaux de neurones profonds (connu sous le nom "deep learning").

L'étude des systèmes biologiques suppose une collaboration pluridisciplinaire ; la bio-informaticienne Alia Ben Kahla et la bio-mathématicienne Ines Abdeljaoued-Tej travaillent sur la combinaison des systèmes polynomiaux biologiques aux techniques de modélisation des données. Annick Valibouze et Ihsen Yengui travaillent sur les plans d'expériences comme application de la théorie de Galois sur les corps finis i.e. sur les corps de Galois. D'autres approches purement statistiques existent et une combinaison des deux outils déterministe (Calcul Formel) et stochastique (construction de graphes ou de réseaux de régulations protéiques) est également à développer. Nous projetons donc une collaboration avec Hatem Hadj Kacem dont la maîtrise des outils de Deep Learning permet d'appréhender le problème de la reconstruction d'un réseau (représentant les interactions entre protéines) sous l'angle de l'apprentissage supervisé. L'objectif est d'approfondir nos connaissances sur les systèmes polynomiaux dynamiques ou galoisiens ainsi que de rechercher des algorithmes efficaces pour leur calcul. Les algorithmes portant sur les systèmes polynomiaux dynamiques (ou discrets) seront appliqués à des données biologiques réelles. La réussite de notre projet de recherche passera par un travail d'équipes pluridisciplinaires : assurer l'intégration des données d'expression des produits de gènes, ajuster des approches de plans d'expérience à la construction de systèmes polynomiaux et de graphes de régulation géniques, ceci passera par un choix judicieux des données.

3 Présentation de chaque équipe engagée

Il y a une équipe française appartenant à Sorbonne Université et deux équipes tunisiennes de l'Université de Tunis El-Manar et de l'Université de Sfax.

3.1 Identification des participants en France

Depuis la fusion de P4 et P6, les participants font partie de Sorbonne Université. Il faut bien définir le **Chef du projet français**, les **Structures de Recherche** et la **Composition de l'équipe**.

Philippe Aubry : MC, LIP6 (labo Info), calcul formel et Nombres Flous (données approchées)

<https://www.lip6.fr/actualite/personnes-fiche.php?ident=P69>

Annick Valibouze : PR, Dr Adj. ISUP (Inst. de Stat. de Paris), LIP6, LPSM (Proba-Stat), calcul formel, plans d'expériences, nombres flous, systèmes dynamiques, apprentissage et réseaux neuronaux

<https://www-apr.lip6.fr/avb/>

Pierre Saurel : MC-HDR (enseigne à l'ESPE), Equipe RATIONALITÉS CONTEMPORAINES ; Math-Info, en copie, apprentissage par renforcement (thèse et HDR en IA), Membre IEEE et VP Chapitre "Computer (C016)" IEEE France.

<http://www.espe-paris.fr/chercheur/saurel>

3.2 Identification des participants en Tunisie

Il y a deux équipes tunisiennes se trouvant l'une à Sfax et l'autre à Tunis. Ici, il faut insérer le **Chef du projet tunisien**, les **Structures de Recherche** et la **Composition de l'équipe**.

3.2.1 Faculté des Sciences de Sfax

Ihsen YENGUI : PR, Département de Maths : calcul formel, plans d'expériences, ...

Hacem Hadj kacem, MA à la Faculté des Sciences Économiques et de Gestion de Sfax (FSEGS), Département Informatique (Computer Science), membre du Laboratoire ReDCAD, Combinatoire, Parallélisme, calcul distribué.

<http://www.redcad.org/members/hatem.hadjkacem/> (- ACM Professional Member-

IEEE Professional Member (Computer Society) (à compléter, modifier ou améliorer car les infos ne sont pas sur la page).

M. Heni Bouhamed, MA : Advanced Technologies for Image and Signal Processing unit (ATISP, ENET'COM), Machine Learning, Deep Learning, Bayesian Networks, Big data, Hadoop. <http://sites.google.com/site/drhenibouhamed/>).

3.2.2 Laboratoire BIMS de l'Institut Pasteur de Tunis

Laboratoire BIMS - Institut Pasteur de Tunis, Directrice : Alia Benkahla, Biologiste, Bioinformatique

BIMS : Bio-informatique, Bio-mathématiques et Bio-Statistique

<http://www.pasteur.tn/>

Ines Abdeljaoued : MA à ESSAI - Université de Carthage et membre du Laboratoire BIMS, calcul formel, bio-math, bio-info, deep-learning, ...

Alia Benkahla, DR Laboratoire BIMS, Biologiste, Bioinformatique, Institut Pasteur de Tunis.

Khadija El Matteli : Entame sa thèse en février à BIMS (sous la direction de Slimane Ben Miled et Ines Abdeljaoued-Tej), ...

3.3 Coopérations antérieures des partenaires

Une collaboration entre les équipes de Tunisie et de France a commencé à propos des systèmes polynomiaux galoisiens. Elle se traduit par la publication de plusieurs articles dont plusieurs signés en commun à propos des systèmes polynomiaux galoisiens en théorie de Galois et l'utilisation du Calcul Formel en biologie.

Nous avons travaillé sur des techniques relevant de l'algèbre, du calcul formel, du calcul scientifique et des récentes avancées du calcul symbolique/numérique auxquelles nous avons contribué [Orange]. Nous poursuivrons nos travaux sur les systèmes biologiques [Handbook] en général et les systèmes polynomiaux et leurs applications en biologie en particulier [Separateurs]. les systèmes polynomiaux galoisiens étant des systèmes possédant une symétrie particulière ; ils ont été introduits par A. Valibouze afin de calculer sans erreur avec les racines des polynômes univariés.

Depuis de nombreuses années, les équipes du réseau français ont un fort taux de collaboration. Par exemple, au travers du projet Galois dirigé par A. Valibouze, auquel I. Abdeljaoued (Tunisie) a également participé.

D'autre part, il y a eu l'organisation conjointe de conférences :

1) MAP'09, « Mathématiques-Algorithmes-Preuves » , Monastir (décembre 2009) organisé par Henri Lombardi et Ihsen Yengui. Plusieurs membres français et tunisiens de ce projet ont participé à ce colloque.

2) la Session parallèle "Calcul formel et Algèbre constructive" a été organisée par I. Yengui et A. Valibouze à l'occasion du Premier Colloque franco-tunisien de Mathématiques (Djerba, 16-20 mars 2009) organisé conjointement par la SMF et la SMT. Les membres de ce projet ont assisté à cette Session parallèle. M. Barkatou, P. Aubry, I. Abdeljaoued-Tej ont été des conférenciers de ce colloque.

3) La deuxième édition du « colloque franco-maghrébin de calcul formel » a été organisée par les deux responsables (coté français et coté tunisien) A. Valibouze et I. Yengui de ce projet du 29 septembre au 04 octobre 2011 aux îles de Kerkennah (Sfax, Tunisie).

4 Données biologiques et médicales

Je mets ce que je sais - A remplir

- Systèmes dynamiques discrets (Abdeljaoued, Benkahla, ..., Valibouze) et application au Cancer

- Plan d'expériences (Yenguy, Valibouze, etc : voir section plus bas)

4.1 Systèmes dynamiques discrets

La Topologie de la carte d'interactions moléculaires (la mesure physique ou les prédictions informatiques des interactions protéiques, des interactions entre ADN et protéines et entre les divers produits de gènes) est codée par un jeu de variables dans lesquelles le temps agit dynamiquement. Le défi serait de mesurer ces changements et extraire de ces mesures toute la dynamique inhérente à ces changements. Ce travail ne peut être entrepris sans l'intégration des données bio-moléculaires et bio-informatiques. Cette intégration et cette collecte est une partie importante de notre travail [Handbook]. Mais d'autres thèmes purement mathématiques peuvent être développés afin de faciliter le traitement des données biologiques : il s'agit par

exemple de l'utilisation du Calcul Formel et plus particulièrement des systèmes polynomiaux.

Un certain nombre d'outils statistiques ont été développés (notamment les réseaux bayésiens dynamiques), mais très peu de travaux portent sur les systèmes polynomiaux biologiques (outre le cas particulier des systèmes polynomiaux booléens). En effet, dès il y a moins d'un demi siècle, les travaux de R. Thomas datent de 1973 [Thomas], l'étude des systèmes dynamiques booléens a connu un formidable développement tout au long des décennies précédentes. Un travail important a été réalisé il y a moins de 10 ans sur ce sujet : l'équipe de R. Laubenbacher a généralisé cette théorie aux systèmes dynamiques polynomiaux, permettant de réaliser des modélisations innovantes en Biologie [Laubenbacher]. Ce que nous entreprenons dans le cadre de ce projet, c'est l'étude des systèmes biologiques qui suppose une combinaison d'outils déterministes (calcul formel) et d'outils stochastiques (construction de graphes ou de réseaux de régulations géniques ou protéiques pour modéliser des systèmes complexes).

Parmi les systèmes dynamiques à modéliser, nous avons entamé une étude à plusieurs échelles de tumeurs cancéreuses solides. Partant d'un nombre limité de cellules souches cancéreuses, notre objectif était de comprendre l'évolution de la taille de la tumeur cancéreuse, sachant que la croissance tumorale est tributaire de l'environnement (nutriments, densité de la matrice extra-cellulaire) mais aussi du génotype des cellules. Ce travail a été réalisé lors du master de Khadija El Matteli (soutenu en Octobre 2017) : outre la synthèse des outils utilisés jusque là, il s'agissait d'étudier les réseaux de neurones modélisant le phénotype, des EDP de réaction-diffusion pour expliquer la distribution du glucose, de l'hydrogène, de l'oxygène et de la densité de la matrice extra-cellulaire. La solution de ces équations influe sur la décision du réseau neuronal. Les simulations informatiques se sont basées sur le modèle d'automate cellulaire pour la croissance tumorale solide, dans lequel chaque cellule est équipée d'un réseau de neurones simulant la réponse micro-environnementale.

Des perspectives de recherche sont proposées, notamment avec l'étude du rôle des gènes sur l'hétérogénéité de la tumeur cancéreuse (selon le phénotype). Nous allons étudier également l'effet des mutations qui se produisent au niveau cellulaire et lors de l'apparition de l'angiogenèse. L'objectif principal de ces modèles de simulation est d'étudier la dynamique de la tumeur et son évolution ; obtenir des résultats statistiques sur le développement de la tumeur et sa croissance dans son environnement ; avoir des informations prédictives sur l'invasion tumorale. Cette modélisation per-

met aussi de combiner des données de plusieurs échelles et de comparer entre les tumeurs bénignes et celles qui sont cancéreuses.

La plupart des simulations se font sur SageMath, qui est un logiciel de mathématiques sous licence GPL. Il combine la puissance de nombreux programmes libres dans une interface commune basée sur le langage de programmation Python. Son objectif est de créer une alternative libre à Magma, Maple, Mathematica et Matlab. SageMath peut être utilisé soit en ligne de commande, soit avec son bloc-note originel ou soit avec le bloc-note Jupyter développé récemment pour piloter la plupart des langages de programmation. Dans [Thiery], nous avons présenté quelques exemples pratiques d'utilisation de SageMath dans la Biologie Théorique. Nous avons utilisé la plateforme cocalc.com pour effectuer nos simulations. Cette plateforme nous permettra effectivement de faire converger nos simulations numériques.

4.2 Plan d'expériences

L'introduction systématique de méthodes statistiques en planification d'expériences est due au statisticien britannique R. A. Fisher. Dans les années 1925-1937, au sein de la station de recherche agronomique de Rothamsted (GB), il entreprit de tenir compte dans les résultats d'expériences de l'hétérogénéité des parcelles, lieux des expérimentations, en y comparant les rendements des différentes variétés de semences qui y étaient cultivées ; il a ainsi introduit les plans en blocs complets, en blocs incomplets, etc. La construction et l'utilisation de ces plans est basée sur trois principes, à savoir la répétition, la randomisation et le contrôle local. La répétition a pour objectif de permettre une estimation de la variabilité résiduelle, et, d'accroître la précision de l'expérience. La randomisation permet d'obtenir des estimations non biaisées de la variabilité résiduelle et de l'influence des variétés sur les rendements. Comme la répétition, le contrôle local a pour but, d'accroître la précision de l'expérience.

Depuis, afin d'étudier un phénomène de nature quelconque : médical, agronomique, physique, chimique, marketing ou autre, les bonnes pratiques consistent à chercher à effectuer, lorsque cela est possible, un nombre limité d'expériences, chacune de celles-ci apportant une information sur le phénomène permettant de l'identifier si possible de manière unique. Le phénomène est souvent régit par des modèles qui peuvent être pré-établis ou non, ce qui nécessite alors de les identifier par des méthodes telles celles d'apprentissage.

Par exemple, pour des expériences à données numériques discrètes dites *qualitatives*,

tels les Plans Factoriels Fractionnaires, PFF, (jumeaux théoriques des codes correcteurs d'erreurs linéaires), le modèle est supposé connu a priori (une pré-étude peut-être nécessaire pour cela). Si un tel modèle est un polynôme, son nombre de variables est alors le nombre de "facteurs" et ses coefficients appartiennent à une structure algébrique contenant les valeurs que peuvent prendre ces facteurs (comme l'anneau des entiers, voir par exemple [Hanan-Tinsson], ou un anneau Z/nZ). Pour une expérience donnée et à l'aide d'un nombre minimal de points expérimentaux, il s'agit de déterminer les paramètres du modèle ; à savoir les coefficients du polynôme dans l'exemple précédent.

Eviter de réaliser toutes les expériences possibles pour déterminer les paramètres d'un modèle est exactement l'objectif de l'étude du plan d'expérience ; en effet, le plus souvent il y a impossibilité dans un temps raisonnable et/ou en pratique de mener toutes les expériences. La question est donc de choisir les points expérimentaux nécessaires et suffisants à "interpoler" le modèle ou tout au moins l'approcher au mieux selon le nombre d'expériences possiblement réalisables en pratique et/ou de leurs précisions ; cela peut inclure la nécessité d'éviter la mise en danger d'un patient ou encore d'un volontaire dans le cas d'essais cliniques. "Interpoler le modèle" signifie déterminer ses paramètres à partir des facteurs, d'où la dénomination "Factoriel" (au sens "interpolation mathématique" ou "régression"). Nous disons qu'il s'agit de déterminer une *fraction* du Plan Factoriel. L'exemple que chacun pourra comprendre est le suivant : s'il n'y a qu'un seul facteur et que le modèle est un polynôme de degré n (d'une variable puisque "un seul facteur") ses coefficients liés à l'expérience sont déterminés par une interpolation de Lagrange classique : il suffit de réaliser l'expérience sur $n + 1$ points distincts quelconques qui ainsi fractionneront le PFF ; l'identifiabilité du modèle dépend donc du nombre de valeurs de points expérimentaux distincts possibles pour l'expérience, à savoir au moins $n + 1$ pour notre exemple.

L'étude des Plans d'Expériences est bien plus étendue que celle essentielle des Plans Factoriels Fractionnaires. Les données (résultats de l'expériences) peuvent être de nature quantitative : réelles (une température, une concentration, une pression, une longueur, ...), scalaires entières (un nombre d'éléments, ...) ou de nature qualitative (un type de matériel, de médicament, le genre des individus considérés, leurs niveaux de diplômes, leurs catégories socio-professionnelles, ...).

4.3 Big-data - Données Massive

Hacem Hadj Kacem et Heni Bouhamed

Voici les données brutes sortie des mails que j'ai transmis

4.3.1 ce qui est proposé (voir mail)

1/ Proposer un formalisme de modélisation permettant de construire un modèle non linéaire et scalable pour des données massives (big data), le Deep Forward Neural Network par exemple, avec des données mixtes,

2/ Produire une implémentation "scalable" permettant de déterminer, après un parcours "exhaustif", le meilleur modèle (polynôme) afin de le comparer avec ceux (modèles) trouvés par le "space-filling",

3/ Produire un algorithme parallèle pour le "space-filling"

4.3.2 Réponses à mes questions (voir mail)

Modèle scalable Big Data : modèle capable de s'adapter lorsqu'il y a un volume énorme de données à traiter ou lorsqu'il y a un nombre de traitement séquentiel énorme (dépassant les capacités des serveurs les plus puissants) à effectuer... (voir « Understanding Big Data Scalability, Cory Isaakson, Prentice Hall, 2014)

1/ Proposer un formalisme de modélisation permettant de construire un modèle non linéaire et scalable pour des données massives (big data), le Deep Forward Neural Network par exemple, avec des données mixtes,

Comment voyez-vous cela ? Cette phrase est trop englobante pour que je comprenne vers quoi vous voulez aller.

n On veut aller vers la construction d'un modèle (Apprentissage d'un réseau de neurones profond) capable de s'adapter lorsqu'on a un volume de données énormes (Big Data) à traiter (une base de données distribuée/répliquée, HDFS, cassandra etc... par exemple).

2/ Produire une implémentation "scalable" permettant de déterminer, après un parcours "exhaustif", le meilleur modèle (polynôme) afin de le comparer avec ceux (modèles) trouvés par le "space-filling",

Vous faite référence à 1/ ici ? Là encore, c'est trop général pour que je comprenne de quoi il s'agit.

n On voulait dire : produire une implémentation exploitant le multiprocessing (distribuer les traitements sur les core des processeurs disponibles), c'est facilement fais-

able avec python et permettra de trouver la meilleure solution avec un parcours exhaustif (même avec des données massives Big Data), qui pourra être comparé par la suite avec les résultats du space-filling (Afin de pouvoir l'évaluer)

3/ Produire un algorithme parallèle pour le "space-filling",

L'aspect matriciel laisse envisager cela.

n On voulait dire : produire une implémentation du « space-filling » exploitant le multiprocessing (distribuer les traitements sur les core des processeurs disponibles), ce qui va permettre de l'adapter (space-filling) pour le traitement des données massives Big Data...

5 Données Floues et Plan d'expériences

Aubry - Valibouze

sorti d'un projet de thèse avec Philippe (il y a des choses inutiles à retirer)

5.1 Données Floues

Les nombres flous interviennent dans de multiples applications dès lors qu'ils s'agit de traiter des données incertaines : actuariat, finance, biologie, chimie, ... (voir [1] et [2]).

Depuis peu, une méthodologie algébrique de résolution est appliquée aux nombres flous de supports bornés. Nous avons appliqué cette méthodologie aux nombres flous triangulaires dans [3] et [4].

L'algébrisation d'un nombre flou est rendue possible car sa définition est formelle (voir, par exemple, les publications de Didier Dubois et Henry Prade) . Un nombre flou se définit comme une courbe continue définie par $(x, N(x))$ sous certaines contraintes. Par exemple, un nombre flou triangulaire est défini par un triplet (n, a, b) de trois valeurs réelles telles que $a \leq n \leq b$, $N(n)=1$, $N(x)$ est nulle en dehors de l'intervalle $]a, b[$, $N(x)$ décrit une droite pour x entre les valeurs a et n (resp. n et b).

La méthodologie de résolution algébrique offre une certification des résultats mais aussi une résolutions globale ce que ne peuvent garantir ou offrir les méthodes locales utilisées avant l'introduction de l'algébrisation (voir [5] à [8]).

Un paquetage Fuzzy du système de Calcul Formel libre SageMath (Langage Python) a été conçu par Jérémy Marrez, M2 Math-Info USVQ, lors de son stage de M2 (financement LIP6 et co-encadrement P. Aubry et A. Valibouze). Jérémy Marrez est actuellement en thèse au LIP6 avec J.C. Barjard (Directeur du LIP6) dans le domaine de la cryptographie.

Nous proposons d'étendre cette méthodologie formelle des nombres flous triangulaires à diverses familles de nombres flous à supports bornés, comme les polynomiaux : les triangulaires sont polynomiaux de degrés 1, les quadratiques de degré 2.

Il s'agira aussi de développer des programmes dans le logiciel SageMath et s'intéresser aux champs d'applications et comparer les résultats aux méthodes locales (réseaux de neurones, par exemple).

5.2 Traitement des données mixtes

Yengui - Valibouze et qui veut bien

C'est de là que c'est parti

Il s'agit d'étudier des plans d'expériences dits "space filling" qui explorent au mieux l'espace des paramètres.

Plus précisément, il est proposé d'étudier une méthode particulière pour la construction d'un plan d'expérience "space-filling" adapté à des variables mixtes ; i.e. des expériences qui regroupent des variables à la fois quantitatives et qualitatives. L'échantillonnage sur hypercube latin (Latin Hypercube Sampling) est une technique couramment utilisée (voir [Sant2003]). Son succès est dû à ses bonnes propriétés de répartition homogène dans l'espace ainsi que de projection unidimensionnelle, i.e. la projection de l'échantillon sur un des axes des variables possède une distribution uniforme sur celui-ci. Cependant les techniques adaptées à la production de plans d'expériences pour variables mixtes avec potentiellement de "meilleures" propriétés de projections sont plus rarement utilisées et étudiées. Il s'agit de pouvoir aborder la partie algébrique (par exemple, sur les corps finis, dits de Galois), avec un regard de statisticien qui possède des compétences en programmation.

Les données expérimentales en trop grand nombre peuvent être ou bien des données stockées ou récupérées par des flux (Big Data) ou alors des données productibles à partir d'un modèle expérimental.

Nous proposons dans un premier l'étude de l'article [Qian2009] et de ceux qui suivent ainsi que l'implémentation de la méthode sous [SageMath]. Cette méthode permet

la construction d'un plan d'expériences "space-filling" pour des variables mixtes. Elle est appelée "sliced space-filling design" par les auteurs de l'article. Cette méthode concerne le calcul sur les corps finis et plus généralement le Calcul Formel (voir [Legrand-Val]).

Nous nous intéressons aussi aux plans dits du "plus proche voisin" consistant à trouver les plans les plus pertinents pour réaliser des tests à intervalles de temps réguliers dans le secteur d'activités. Par exemples, dans le contexte de la santé, il s'agit d'administrer des traitements à des patients ou à un même patient. Les conditions optimales de deux structures de corrélation (modèles) ont été déjà étudiées dans le cas général ; à savoir les structures NN_m et $AR(m)$, où l'entier $m > 0$ est la distance temporelle entre deux traitements (voir [Kone-Val1], [Kone-Val2]).

Un autre point d'étude concerne les PFF construits à partir des méthodes de résolution algébrique dont l'article fondateur est dû à [Pistone et Wynn]. Les applications sont nombreuses et variées (voir, par exemple, [Hanan-Tinsson]).

6 Planification des résultats attendus

Ici, insérer la planification des résultats attendus, pour l'équipe tunisienne et l'équipe française : **Intérêts scientifiques, Retombées socio-économiques, Activités de Recherche diplômante**, etc.

6.1 Equipe tunisienne

Avec l'avancement des travaux en cours, d'autres doctorants pourraient rejoindre les équipes dans le cadre d'une co-encadrement.

6.1.1 Intérêts scientifiques

Bénéficier de l'aide des experts français dans la modélisation mathématiques, la conception et implémentation d'algorithmes en vue de les appliquer en traitement du signal et en biologie.

Renforcer la collaboration Calcul Formel/Bio (Tunisie) par des spécialistes français de résolution de systèmes polynomiaux et l'enrichir avec la statistique.

6.1.2 Retombées socio-économiques

L'algèbre a vu son contenu et ses méthodes évoluer beaucoup récemment grâce à des progrès conceptuels et logiciels qui rendent possibles de multiples applications, notamment dans le domaine de la protection de l'information, du traitement du signal et de la biologie.

6.2 Equipe française

6.2.1 Intérêts scientifiques

- Proposer des sujets de thèses en rapport avec les thèmes de recherche des équipes françaises. - Renforcer les collaborations existantes qui ont déjà donné lieu à des publications (voir listes de publications). - Faire collaborer simultanément des spécialistes des 3 domaines Statistique, Biologie et Calcul Formel ; ces collaborations n'étant actuellement que bi-thématiques Stat/Bio (France) et Calcul Formel/Bio (Tunisie) en bénéficiant des compétences acquises par l'équipe tunisienne. - Trouver de nouvelles applications aux systèmes polynomiaux et faire avancer la recherche fondamentale face à de nouvelles problématiques.

6.2.2 Retombées socio-économiques

Faire participer des jeunes chercheurs dans les projets de recherche.
Que les nouvelles approches Sta/Bio, Calcul-Formel/Bio mises en synergies apportent des solutions nouvelles en terme de logiciels et avancées des connaissances dans les systèmes dynamiques biologiques ; nous espérons tout particulièrement un impact important dans le domaine médical.

7 Besoins en matériel/séjours

1 ordinateur portable (1 To, Intel Core i7, Quad-core, 15 pouces) ;
1 carte graphique pour effectuer des calculs sur le serveur de BIMS
1 séjour (junior) en France (pour assister à une école) ;
1 séjour (junior) pour un colloque ou une conférence ;

Références

- Broniatowski-Celant , M. Broniatowski et G. Celant, **Interpolation and Extrapolation Optimal Designs 1**. Wiley, 2016, Print ISBN: 9781848219953, Published Online: 1 APR 2016, Online ISBN: 9781119292272, DOI: 10.1002/9781119292272.
- Hanen-Tinsson Hanna Hanen, Walter Tinsson. Plans d'expérience pour mélange de mélanges. 41-ièmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <https://hal.inria.fr/inria-00386588/file/p34.pdf>, HAL-ID = <inria-00386588>.
- Kone-Val1 M. Koné, A. Valibouze, Plans en blocs incomplets : application à la santé. Proceedings des Journées Internationales Analyse Statistique : Théorie et Applications, JIASTA2012, Oujda, Maroc, 113-116, 2012.
- Kone-Val2 M. Koné, A. Valibouze, Universal optimality of nearest neighbour balanced block designs using any order correlated models. Preprint, 2016.
- Legrand-Val R. Legrand, A. Valibouze, **Calcul Formel avec Maple**. (Livre interactif) Science Active Peter Burch (Ed.), Paris, France, ISBN : 978-2-917398-03-6, 2008.
- Pistone et Wynn Pistone G., Wynn H., Generalised confounding with Gröbner bases. *Biometrika*, 83, 653-666, 1996.
- Qian2009 P.Z.G. Qian, C.F.J. Wu, Sliced space-filling design. *Biometrika*, 96:4, 945-956, 2009.
- SageMath SageMath, Système de Calcul Formel libre de langage Python interfacé avec R et des dizaines d'autres logiciels libres : <http://www.sagemath.org/fr/> Ouverture de compte libre sur le cloud de SageMath. <https://www.sagemathcloud.com/> (accès libre limité en espace et en temps mais suffisant pour le stage).
- Sant2003 T.J. Santner, B.J. Williams, W.I. Notz, *The Design and Analysis of Computer Experiments*. Springer-Verlag New York, 2003.
- Val A. Valibouze, La théorie de Galois en Informatique. Bicentenaire Evariste Galois. 2008. <https://www-apr.lip6.fr/avb/Bicentenaire.html>.
- Thierry J. Thiéry, I. Abdeljaoued-Tej, K. El-Matteli and S. BenMiled, SageMath pour l'enseignement de la biologie théorique, Juin 26-29, 2017. 37ème Colloque de la Société Francophone de Biologie Théorique.

- Orange I. Abdeljaoued-Tej, S. Orange, G. Renault, A. Valibouze. 2004. Computation of the decomposition group of a triangular ideal, *Journal of Applicable Algebra in Engineering, Communication and Computing*, numéro 15, pages 279-294.
- Handbook L. Guizani-Tabbane, A. Ben Kahla, I. Abdeljaoued-Tej, S. Ben Miled and K. Delagi. 2008. Chapter System biology and infectious diseases, *Handbook of research on systems biology application in medicine*, Max Planck Institute, Berlin.
- Separateurs I. Abdeljaoued-Tej, A. Ben Kahla, G. Haddad, A. Valibouze, A Linear Algorithm For Computing Polynomial Dynamical Biological Systems, soumis à NGC.
- Thomas R. Thomas, Kinetic logic: a boolean approach to the analysis of complex regulatory systems, *Lecture Notes in Biomathematics*, 29, (1979), 507 pp.
- Laubenbacher R. Laubenbacher and B. Stigler, A computational algebra approach to the reverse engineering of gene regulatory networks, *Journal of Theoretical Biology*, 229 :523, 2004.
- 1 First Course on Fuzzy Theory and Applications, Lee, K., *Advances in Intelligent and Soft Computing*, Springer, 2005
 - 2 Fuzzy Systems in Economy and Engineering, J.G. Aluja, A.P. Tacu, H.N. Teodorescu, Publishing House of The Romanian Acad, 1994
 - 3 Finding solutions of fuzzy polynomial equation systems by an Algebraic method. M. Boroujeni, A. Basiri, S. Rahmany, Annick Valibouze *Journal of Intelligent & Fuzzy System* 30, 781-800, 2016
 - 4 Solving Fuzzy Systems in Dual Form Using Wu's Method. Marziyeh Boroujeni, Abdolali Basiri, Sajjad Rahmany, Annick Valibouze *International Journal of Fuzzy Systems (issn=1562-2479)*, Ed. Springer Berlin Heidelberg, 1-11, 2015
 - 5 Newton's method for solving fuzzy nonlinear equations S. Abbasbandy, B. Asady *Applied Mathematics and Computation*, Vol. 159 (2), 349-356, 2004.
 - 6 Numerical solution of fuzzy polynomials by fuzzy neural network. S. Abbasbandy, M. Otadi *Applied Mathematics and Computation*, Vol. 181 (2), 1084-1089, 2006.

- 7 Numerical solution of a system of fuzzy polynomials by fuzzy neural network S. Abbasbandy, M. Otadi, M. Mosleh Inf. Sci., vol. 1788, 1948-1960, 2008.
- 8 Solving fuzzy equations using evolutionary algorithms and neural nets J. Buckley, T. Feuring Y. Hayashi, Soft Comput., Vol. 6 (2), 116-123, 2002.