

## Propositions de Sujets

Professeur référent : Paola Goatin

Mail : [paola.goatin@inria.fr](mailto:paola.goatin@inria.fr)

### Représentation et analyse de données de trafic

Ce projet propose de manipuler et étudier une base publique de données de trafic routier provenant de boucles de comptage des routes de Minneapolis : <http://data.dor.state.mn.us/datatools/>

Après avoir sélectionné et extrait les données correspondantes à des localisations et des plages horaires données [1], on s'attachera à produire et analyser les graphiques des séries temporelles et des diagrammes fondamentaux, en suivant [2, Chapitre 4]. Un soin particulier sera dédié à distinguer et comparer les dynamiques dans les différentes voies de circulations.

On pourra ensuite employer des méthodes statistiques connues pour extraire des informations sur la circulation routière dans le secteur étudié, comme par exemple les horaires de pointe, les caractéristiques communes sur les différents jours de la semaine, des événements extrêmes, etc.

Les références ci-dessous et tout autre document nécessaire seront fournis.

#### Références

- [1] | S. Fan, Mi. Herty, B. Seibold. Comparative model accuracy of a data-fitted generalized Aw-Rascle-Zhang model. *Networks Heterog. Media* 9(2) : 239-268 (2014).
- [2] | M. Treiber, A. Kesting. "Traffic flow dynamics." *Traffic Flow Dynamics : Data, Models and Simulation*, Springer-Verlag Berlin Heidelberg (2013).

Professeur référent : Frédéric Precioso

Mail : frederic.precioso@unice.fr

## Face detection - orientation - recognition

One of the key features when analyzing video content (such as movies for instance) is the detection of faces because we are naturally attracted towards other humans.

The real challenge in this context is to be able to detect face with high precision in real-time to be able to extract facial features which will allow us to capture the orientation of the face. From this point we will try to identify who is talking to who in tv series or movies during dialogue scenes. In parallel to this task we will track the faces detected in the first frame and follow them through the whole scene. From all the face tracks we will build a face recognition algorithm so that we will be able to identify who are the persons in the video. Finally we will learn to recognize the persons (for instance famous actors) in the face sets.

We will apply all these to different contexts:

- Tv series such as friends or Buffy the vampire slayer
- Frontal webcam to determine where the user is looking on the screen
- Videos from inside the tram in Nice in order to check how many passengers are looking at those screens

We will mainly focus on deep learning techniques but we will also consider other machine learning algorithms to get the best and fastest accurate results.

The goals of this project are:

- Detecting face and extract facial features to define face orientation
- Track faces along the video
- Recognize faces by naming actors

### **Pre-requisites:**

For this project we require the students to know how to code in Python and to have basic knowledge in Machine Learning.

### **References:**

- Guo, J., Zhu, X., Yang, Y., Yang, F., Lei, Z., Li, S. Z. (2020). Towards fast, accurate and stable 3d dense face alignment. ECCV 2020.  
[https://github.com/cleardusk/3DDFA\\_V2](https://github.com/cleardusk/3DDFA_V2)
- Ramos, L., Morales, B. (2020). SwiftFace: Real-Time Face Detection. arXiv preprint arXiv:2009.13743.
- Zhu, B., Lin, C., Wang, Q., Liao, R., Qian, C. (2020). Fast and Accurate: Structure Coherence Component for Face Alignment. arXiv preprint arXiv:2006.11697.  
[https://github.com/ChanChiChoi/awesome-Face\\_Recognition](https://github.com/ChanChiChoi/awesome-Face_Recognition)

Professeur référent : Melissa Sanabria  
Mail : sanabria@unice.fr

## Audio and Video Similarity

Analyzing video content to produce summaries and extracting highlights has been challenging for decades. However, the interest for these challenges has lately increased with a focus on event content (cultural events such as Oscars, Cesars, Grammys or Emmys; and sport events). Recently several methods try to solve the problem of video summarization for events however, owing to copyright regulations or lack of standardization, there are not available benchmark datasets which makes more difficult the comparison among techniques.

In order to build a video summarization dataset, we need three important components: the video of the entire event, the video of the summary and the summary intervals. We need to be able to identify as fast as possible (close to real-time) where each video segment found in summary is located in the entire video: what is the frame in the entire video where a video segment extracted from the summary starts and at what frame it ends. In the project, we will focus on videos of soccer games.

One advantage of the use of a popular sport like soccer is that the videos of the games and summaries are available in many platforms. But, even if the summary videos are extracted with an edition program directly from the original video game, the time intervals are not available.

Although there are techniques to find similarities between videos, they are based on frame features which is not the most efficient way for soccer since the images of the field during the entire match are very much alike. However, it is important to remember that a video is composed not only by the images but also by the audio signal. Therefore matching video shots can be found with the help of the two modalities.

The goals of this project are:

- Detect the boundary of the video shots of the summary videos
- Look for the matching shots in the original game video

### **Pre-requisites:**

For this project we require the students to know how to code in Python and to have basic knowledge in Machine Learning.

### **References:**

- Kordopatis-Zilos, G., Papadopoulos, S., Patras, O., Kompatsiaris, I. (2019). Visol: Fine-grained spatio-temporal video similarity learning. In Proceedings of the IEEE International Conference on Computer Vision (pp. 6351 - 6360)
- Avgoustinakis, P., Kordopatis-Zilos, G., Papadopoulos, S., Symeonidis, A. L., Kompatsiaris, I. (2020). Audio-based Near-Duplicate Video Retrieval with Audio Similarity Learning. arXiv preprint arXiv:2010.08737

- Worldveil, “Audio fingerprinting and recognition in python”, <https://github.com/worldvei/dejavu>, 2013
- Wang, A. (2006). The Shazam music recognition service. *Communications of the ACM*, 49(8), 4 - 48.  
<https://github.com/MKLab-ITI/visil>

Professeur référent : Agustin Yabo

Mail : agustinyabo@gmail.com

## **Mathematical modeling and numerical analysis of dynamical allocation of cellular resources in bacteria**

In nature, microorganisms continuously face environmental changes, and thus they have evolved to rapidly adapt their physiology to deal with this phenomenon. This is achieved by dynamically allocating their internal resources to different cellular functions. In order to study their behaviors, such microorganisms can be represented by mathematical models. In particular, simple non-linear systems of ordinary differential equations are highly capable of representing the main cellular functions such as bacterial growth, metabolism and cell maintenance. In the specific case of *Escherichia coli*, a very simple microorganism, studies have shown that bacterial populations seek to maximize their bacterial volume, and so they achieve exponential growth speed in order to survive in changing environments. The field of "Systems Biology" aims to understand some of these biological phenomena using mathematical modeling and numerical simulations. This project consists of proposing biologically relevant variations of the mathematical models of bacteria existing in the research literature, and to aim to understand their naturally-evolved behaviors using optimization techniques. The work will involve analytical approaches to the problem, as well as numerical simulations of different scenarios using Python and Bocop. These kinds of studies can not only help us understand the natural bacterial intracellular mechanisms, but also help improve industrial processes, as well as provide guidance in biotechnological research.

Professeur référent : Guilia Marchello

Mail : guilia.marchello@inria.fr

## Latent Block Model with SEM Inference algorithm

In many applications, it is now frequent to have to summarize large matrices. For instance, the growing number of people using tools like Tripadvisor, Amazon, eBay led the major e-commerce companies to manage massive amounts of data to be able to offer the most suitable products to each of their customers. It is of interest for those companies to cluster both customers and products to better understand purchasing behaviours. The simultaneous clustering of rows and columns of matrices is known in literature as a co-clustering problem. The latent block model (LBM) is often used as a basis for many model-based methods for co-clustering. This model was proposed for the first time by Govaert and Nadif (Govaert G. a., 2003) and it is based on the assumption that rows and columns are grouped in hidden clusters. The model has been extended allowing to deal with counting data (Govaert G. a., 2010), real data (Lomet, 2012), categorical data (Keribin, 2015), ordinal (Jacques, 2018); (Corneli, 2020), functional data (Bouveyron, 2018) and textual data (Bergé et al., 2019). Several inference procedures have been proposed for LBM, among which the most common one is the VEM algorithm. The VEM algorithm often provides satisfactory results but it highly depends on its initial values and it has the tendency to give empty clusters (Christine Keribin, 2015). To overcome these limitations we propose here to analyze the LBM using the SEMGibbs algorithm for model inference and to make use of the ICL criterion for model selection. In fact, while VEM is based on a variational approximation, SEM ? Gibbs is based on a Gibbs sampler to simulate the unknown labels with their conditional distribution knowing the observations and the current values of the parameter set. The method will be tested on simulated dataset and then the model performances will be evaluated on a real dataset.

### Pre-requisites:

machine learning, statistical inference, clustering, classification

### References:

- Govaert, G. a. (2010). Latent block model for contingency table. *Communications in Statistics - Theory and Methods*, 416–425
- Keribin, C. a. (2015). Estimation and selection for the latent block model on categorical data. *Statistics and Computing*, 1201–1216
- Jacques, J. a. (2018). Model-based co-clustering for ordinal data. *Computational Statistics & Data Analysis*, 101–115
- Corneli, M. a. (2020). Co-clustering of ordinal data via latent continuous random variables and not missing at random entries. *Journal of Computational and Graphical Statistics*, 1–15
- Bouveyron, C. a.-X. (2018). The functional latent block model for the co-clustering of electricity consumption curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 897–915
- Govaert, G. a. (2003). Clustering with block mixture models. *Pattern Recognition*, 463–473

- Lomet, A. (2012). Un modèle de mélange pour la classification croisée d'un tableau de données continue. Compiègne
- Christine Keribin, V. B. (2015). Estimation and Selection for the Latent Block Model on Categorical Data. Statistics and Computing, 1201-1216

## Semi-Supervised Learning

Semi-supervised learning is a branch of machine learning that aims to combine both supervised learning and unsupervised learning. Typically, semi-supervised learning algorithms attempt to improve performance in a classification problem by leveraging additional data points for which the label is unknown and might be used to aid in the classification process [3]. Indeed, labeled examples are often scarce in certain domains such as computer-aided diagnosis or segmentation. In those cases, it may be difficult to construct a reliable supervised classifier. Semi-supervised learning frameworks have received a considerable amount of attention during the past two decades. Most of the proposed approaches can be classified as “inductive” methods remain (attend to find a classification model) but some really interesting “transductive” methods remain (obtaining label predictions for the given unlabelled data points). See [3] for the complete taxonomy of Semi-supervised framework.

In practice, in [2], the authors proposed a graph-based method : labeled and unlabeled data are represented as vertices in a weighted graph, the learning problem is then formulated in terms of a Gaussian random field on this graph. The solution of such a problem is then characterized by harmonic functions over the graph.

### Project:

The student should start to read and clearly understand the paper [2]. Then, the first step would be to implement the proposed methods and try to reproduce the results of the paper. In [1], the authors proposed a standardized pipeline to evaluate inductive methods. The students should try to adapt this pipeline to evaluate our transductive method on nwdays datasets.

### Going Further:

Due to their time complexity, transductive approaches do not scale to large amounts of data ([3]). The students may try to overcome this issue by exploiting the recent advances in the task of representation learning.

### References:

- Oliver et al. Realistic evaluation of deep semi-supervised learning algorithms. NeurIPS, 2019. <https://arxiv.org/pdf/1804.09170.pdf>
- Zhu et al. Semi-supervised learning using gaussian fields and harmonic functions. ICML, 2002. <http://mlg.eng.cam.ac.uk/zoubin/papers/zgl.pdf>
- Jesper E. van Engelen and Holger H. Hoos. A survey on semi-supervised learning. Machine Learning, 2018. <https://link.springer.com/article/10.1007/s10994-019-05855-6>



Professeur référent : Lucile Sassatelli

Mail : sassatelli@i3s.unice.fr

## Gravitational Laws of Attention in 360° Videos

Virtual Reality (VR) is growing fast with accessible head-mounted displays (like Oculus Rift or Google Cardboard). The attentional process in VR, that is, where the user moves and watches over a 360° sphere, is not well known. It is however crucial to better understand and predict the user's behavior for several types of applications. A few works have tackled this problem so far, using saliency models and deep neural networks [1,2] to predict the user's head and gaze positions, as depicted above. Recently, Zanca et al. proposed a set of gravitational laws to predict the motion of the human gaze watching a regular video, obtaining compelling results [3].

### Objective:

The goal is to test a gravitational model to predict the human motion in 360° videos.

Phases:

- First the students will familiarize with article [3], and with our new work modeling the head motion with angular momentum and torque, undergoing forces created by attracting areas of the 360° video. In particular, the existing code will be understood to be extended.
- Then, the students will code components including visual features (objects location, amount of motion, lower-level features), and reference metrics to assess the model on established datasets.
- Finally, performance results of the gravitational model will be obtained and analyzed.

### References:

- [1] | Marc Assens Reina, Xavier Giró-i-Nieto, Kevin McGuinness, Noel E. O'Connor. SaltiNet: Scan-Path Prediction on 360 Degree Images Using Saliency Volumes. The IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2331-2338
- [2] | Marc Assens Reina, Xavier Giró-i-Nieto, Kevin McGuinness, Noel E. O'Connor. Scanpath and saliency prediction on 360 degree images. Elsevier Signal Processing: Image Communication. Volume 69, November 2018, Pages 8-14.
- [3] | D. Zanca, S. Melacci and M. Gori, "Gravitational Laws of Focus of Attention," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019. doi: 10.1109/TPAMI.2019.2920636.

## Autour de la cicatrisation in silico

Une méthode largement utilisée pour tester la capacité de réorganisation d'un tissu épithélial est le test de cicatrisation (ou wound healing) : simple à mettre en place et ne nécessitant pas d'appareillage complexe. Mais leur analyse précise et la possibilité d'extraire des informations fiables l'est beaucoup moins. Les paramètres extraits (manuellement) sont souvent relatifs, utilisateur dépendant, et leur exploitation lacunaire. D'où la nécessité d'une modélisation fine, rendant compte de la réelle dynamique biologique du tissu aussi fidèlement que possible. Dans cet objectif, un des modèles les plus populaires est celui F-KPP développé indépendamment par Fisher et par Kolmogorov-Petrovsky-Piscounov. C'est une équation de diffusion-réaction, qui permet de modéliser la propagation de front d'onde formé par les bords de cicatrice. Nous proposons ici une étude en 3 volets, qui peut être partiellement ou totalement collaborative (1, 2 ou 3 groupes), sur trois aspects importants de la modélisation et simulation de dynamique de cicatrisation :

1. Calibration et validation du modèle classique F-KPP utilisant des vidéoscopies biologiques réelles
2. Etude d'une adaptation du modèle classique aux cicatrises inhibées et activées (par des protéines exogènes)
3. Comparaison du modèle continu de F-KPP à une simulation de type agent based modeling

Les 3 variantes du projet nécessitent uniquement la connaissance de la méthode des différences finies (1D et/ou 2D). Elles permettent aux élèves d'acquérir de bonnes pratiques en implémentation D.F. en 2D, et de se familiariser avec de nouveaux modèles en biologie mathématique. Le développement sur [gitlab.polytech.unice.fr](https://gitlab.polytech.unice.fr) et le recours à la machine Abel sont à privilégier. On trouvera page suivante le détail de chacun de ces trois volets.

### Sujet 2.1 : Modèle F-KPP classique : calibration et validation

Le modèle classique de F-KPP est une équation de la chaleur à coefficients constants, avec un terme de réaction non linéaire de la forme  $u(1 - u)$  où  $u$  est la densité des cellules. Avec un petit coefficient de diffusion, ce modèle permet de "fermer" les cicatrices à vitesse constante (qui dépend des paramètres de l'équation). Nous disposons de séquences vidéos de cicatrises de cellules MDCK (cellules de rein de chien, très populaires en étude de dynamique cellulaire chez les biologistes) sur lesquelles nous nous baserons pour identifier au mieux les paramètres du modèle mathématique. Il s'agit donc, dans ce volet, de comprendre à minima le problème biologique de cicatrises, d'implémenter l'équation de F-KPP en 2D par différences finies, et de faire un tout petit peu de traitement d'images pour extraire les aires de blessure à partir des images vidéo. On sera satisfait si, notamment, on montre que les vitesses de cicatrises du modèle F-KPP et celle de

l'expérience biologique sont proches (pour un jeu de paramètres du modèle bien trouvé).

## **Sujet 2.2 : Modèle de Fisher à coefficients variables en cicatrisation inhibée/activée**

Le but de ce volet du projet est d'étudier un modèle de cicatrisation 2D en situation de dynamique activée ou inhibée. Dans un premier temps, on considère l'équation dite de Fisher à coefficients constants que l'on implémentera en différences finies. Ensuite, on souhaite modifier les coefficients de cette équation (de constants, ils deviennent sigmoïdes) de sorte à modéliser l'accélération ou l'inhibition de la cicatrisation. Un but important du projet est de valider l'hypothèse selon laquelle des coefficients sigmoïdes y parviendraient.

## **Sujet 2.3 : Multi-agent versus le modèle de F-KPP**

Nous disposons de deux approches pour modéliser l'interaction entre cellules dans un processus biologique donné, une continue (équations aux dérivées partielles) et l'autre discrète (individu based modeling) basée sur une description comportementale action/réaction dans les systèmes multi-agent. Le but du projet est de mener une étude comparative des deux approches (évolution d'aire de la blessure, vitesse du front cellulaire, etc), en s'initiant à la librairie python Mesa dédiée aux systèmes multi-agent, en implémentant un modèle simple de cellules en interaction, et en menant une ou deux simulations sur des cas-tests de cicatrisation fournis par les biologistes.

## Approximation softmax en machine learning

La classification supervisée cherche à attribuer une classe, parmi K classes possibles (par exemple une image en entrée du classifieur, les classes peuvent être du type “chien”, “chat”, “oiseau”, etc), à une donnée en entrée du classifieur (image, texte, vecteur numérique, etc). Dans le domaine de l'apprentissage profond, la fonction “softmax”[1] est une fonction universellement utilisée pour transformer un vecteur de nombres réels  $z$  en une loi de probabilité discrète  $p(z)$  :

$$p(z) = (p_1, p_2, \dots, p_K) = \text{softmax}(z) = \frac{1}{\sum_{l=1}^K e^{z_l}} (e^{z_1}, e^{z_2}, \dots, e^{z_K}, \forall z = (z_1, z_2, \dots, z_K) \in \mathbb{R}^K)$$

Ce type de transformation est indispensable dans les problèmes de classifications supervisées afin de choisir parmi les K classes possibles : la valeur  $p_i$  correspond typiquement à la probabilité de choisir la classe  $i$ . La fonction “softmax” est également appelée la fonction “sofrargmax” dans la mesure où elle peut être interprétée comme l'approximation continument différentiable de la fonction “one-hot argmax” qui sélectionne l'élément maximum dans un vecteur. Plus précisément, l'encodage “one hot” [2], très utilisé en machine learning, consiste à coder l'entier identifiant la classe  $i$  sous la forme d'un vecteur binaire de dimension K composé uniquement de 0 sauf pour la position  $i$  qui est indiqué avec le chiffre 1. Par exemple, pour K=3, les trois classes sont codées sous la forme  $h(1) = (1, 0, 0)$ ,  $h(2) = (0, 1, 0)$  et  $h(3) = (0, 0, 1)$  où  $h(i)$  est l'encodage “one-hot” du chiffre  $i$ . Avec cette notation, la fonction “one-hot argmax”[1] est définie par

$$h(k) = \text{argmax}(z) = \text{argmax}(z_1, z_2, \dots, z_K) \text{ quand } z_k = \max_{1 \leq i \leq K} z_i$$

La fonction « one-hot argmax » est la fonction utilisée par le classifieur de Bayes que le réseau de neurones souhaite systématiquement reproduire. Le but de ce projet est d'étudier la précision de l'approximation « softmax ». En utilisant un jeu de données simulé avec plusieurs classes à reconnaître, les étudiants devront comparer les performances d'un classifieur de Bayes avec la fonction « one-hot argmax » à un classifieur de Bayes avec « softmax ». Ils devront notamment calculer le risque d'erreur global du classifieur dans chaque cas. Dans ce but, les étudiants utiliseront Python, et notamment la différentiation automatique [3,4], pour effectuer cette comparaison. Ils pourront notamment déterminer les proportions de données par classe qui maximise le risque du classifieur « softmax ».

### Objectifs :

Le projet est composé de trois tâches principales :

- La première tâche consiste à se familiariser avec le classifieur de Bayes et la fonction “one-hot argmax”
- La seconde tâche consiste à étudier le classifieur de Bayes avec « softmax »
- Enfin, les étudiants doivent comparer les deux approches, notamment en fonction des proportions des données d'apprentissage pour chaque classe

Les développements informatiques et l'analyse mathématique seront présentés dans un fichier Jupyter Notebook et un rapport technique.

**Références :**

[1 ] [https://en.wikipedia.org/wiki/Softmax\\_function](https://en.wikipedia.org/wiki/Softmax_function)

[2 ] <https://en.wikipedia.org/wiki/One-hot>

[3 ] [https://www.cs.toronto.edu/~rgrosse/courses/csc321\\_2018/slides/lec10.pdf](https://www.cs.toronto.edu/~rgrosse/courses/csc321_2018/slides/lec10.pdf)

[4 ] <https://github.com/HIPS/autograd>

Professeur référent : Lionel Fillatre  
Mail : lionel.fillatre@i3s.unice.fr

## Réseau de neurones pour résoudre les équations différentielles

Les équations différentielles occupent une place très importante dans de nombreux domaines de l'ingénierie [1]. Les réseaux de neurones sont devenus une technique incontournable dans le domaine du machine learning [2]. Récemment, il y a eu un rapprochement significatif entre ces deux disciplines : utiliser un réseau de neurones pour modéliser la solution d'une équation différentielle, puis calculer cette solution [3]. Le but de ce projet est de résoudre une équation différentielle ordinaire (linéaire ou non-linéaire) avec un réseau de neurones. Nous ne considérerons que des équations différentielles pour lesquelles nous disposons d'une solution analytique connue [1]. Les étudiants devront comparer la solution numérique fournie par le réseau de neurones avec la solution exacte. Ils devront également comparer cette solution avec celle fournie par des approches numériques classiques (méthode d'Euler par exemple [1]). Nous utiliserons Python, et notamment l'environnement Pytorch [4] dédiée au réseau de neurones profonds, pour réaliser ce projet.

### Objectifs :

Le projet est composé de trois tâches principales :

- La première tâche consiste à se familiariser avec les réseaux de neurones profonds et Pytorch. De nombreux documents et codes sont disponibles en ligne [4]
- La seconde tâche consiste à concevoir et programmer un réseau de neurones profonds pour résoudre une équation différentielle ordinaire connue [1, 3].
- Enfin, les étudiants doivent résoudre une équation différentielle avec une méthode classique et comparer cette approche classique avec l'approche à base de réseaux de neurones [1].

Les développements informatiques et l'analyse mathématique seront présentés dans un fichier Jupyter Notebook et un rapport technique.

### Références :

- [1 ] [http://faccanoni.univ-tln.fr/user/enseignements/20152016/R33-R43\\_L2.pdf](http://faccanoni.univ-tln.fr/user/enseignements/20152016/R33-R43_L2.pdf)
- [2 ] <http://neuralnetworksanddeeplearning.com/>
- [3 ] <https://maziarraissi.github.io/PINNs/>
- [4 ] <https://pytorch.org>

Professeur référent : Pierre Alliez

Mail : pierre.alliez@inria.fr

## Régularisation de formes 3D par dilatation/érosion

L'opération consistant à dilater un solide d'un rayon  $r$  puis à dilater son complémentaire d'un rayon  $d < r$  produit, dans certain cas, un nouveau solide avec un bord lisse. Ce principe a été largement utilisé en modélisation géométrique ou en traitement d'images, et une condition robuste et optimale a été démontrée pour garantir la régularité de l'objet obtenu par dilatation et érosion [1].

### Objectif

L'objectif de ce projet consiste à concevoir un algorithme d'évaluation d'une fonction implicite du double offset, à partir d'un maillage surfacique triangulaire. Plus précisément, on cherche à évaluer efficacement une fonction de  $\mathbb{R}^3$  dans  $\mathbb{R}$  permettant d'évaluer, en tous points de l'espace, la fonction distance signée à la surface du double offset. Le premier offset (positif) est simple à évaluer avec les structures accélératrices de la bibliothèque CGAL ([http:// www.cgal.org](http://www.cgal.org)). Le second offset, quant à lui, requiert de concevoir une discrétisation à l'aide de quadratures adaptatives, et de recourir à une interpolation numérique. On cherchera à concevoir une méthode frugale en mémoire et une mise en oeuvre en C++ comme un « oracle » implicite de l'algorithme de maillage 3D de CGAL [2].

### Compétences requises

Mathématiques appliquées, calcul géométrique, C++, programmation générique.

**Mots-clés** : fonction distance, régularisation, fonction distance, axe médian, C++, CGAL.

### Références

- [1] | Shape Smoothing using Double Offsets. Frederic Chazal, David Cohen-Steiner, Andre Lieutier, Boris Thibert. In Proceedings of SPM'07.
- [2] | CGAL 3D Mesh Generation  
[https://doc.cgal.org/latest/Mesh\\_3/index.html#Chapter\\_3D\\_Mesh\\_Generation](https://doc.cgal.org/latest/Mesh_3/index.html#Chapter_3D_Mesh_Generation)

Professeur référent : Florent Lafarge

Mail : florent.lafarge@inria.fr

## Détection de surfaces planaires et cylindriques dans un nuage de points orientés

Le but de ce mini-projet est de détecter des surfaces planaires et cylindriques dans des nuages de points orientés, c'est-à-dire où chaque point est associé à un vecteur indiquant le vecteur normal estimé de la surface en ce point. Ces nuages de points orientés, générés à partir de capteurs laser ou multi-vue stéréo de scènes physiques, seront fournis comme entrées du problème. Le mini-projet consiste à implémenter en C/C++ une fonction permettant de détecter et extraire des surfaces planaires par une méthode d'accumulation sur une sphère de Gauss vue durant le cours d'Interpolation numérique. La fonction devra dépendre de 2 paramètres : un paramètre fixant le pas de discrétisation de la sphère de Gauss, et un paramètre spécifiant la sensibilité de détection des plans et de cylindres. Outre les supports du cours et les interactions avec l'encadrant, les étudiants pourront également s'appuyer sur la partie 4 de la publication suivante : [http://qianyi.info/docs/papers/eccv14\\_pipe-run.pdf](http://qianyi.info/docs/papers/eccv14_pipe-run.pdf).

### Taches à réaliser

- Installer des bibliothèques et des logiciels pour l'édition et la compilation de codes C++
- Prendre en main un code C++ « modèle » fourni par l'encadrant
- Développer une structure de données permettant de stocker l'accumulation d'hypothèses sur une sphère de Gauss
- Développer une fonction permettant la détection de plans
- Développer une fonction permettant la détection de cylindres



## Matrix vs. Tensor Image Compression

Grayscale images in their raw format can be seen as matrices of data. By evaluating their low rank approximation through the singular value decomposition (SVD), one can simply store their truncated SVD factors instead of the original matrix itself. If the image can be approximated well by a matrix of sufficiently low rank, this modified storage process may lead to a significant compression of the image representation. Instead of carrying out the low rank approximation of the image itself, one can split the image in many small patches and apply the low rank approximation of the matrix formed by the stacked vectorized patches. This approach may achieve better results than the previous one, whenever the global image can not be well represented by a low rank matrix but its local behavior can be well represented by linear combinations of a few shared patterns. As a third approach, instead of vectorizing the patches and storing them into a matrix, one can store them directly in a data cube called a third order tensor. Similarly as for matrices, tensors also have ranks and low rank approximations [1]. One such low rank approximation is a direct extension of the SVD called the multilinear singular value decomposition (MLSVD) [2]. The low rank approximation of the patches tensor has been used for example in image denoising [3], in this project we want to test it for image compression, by keeping only the truncated factors obtained as a result of the MLSVD. The students will code in python the three compression approaches above and compare their performances using a test image data base, for example [4]. Further developments of the project may include how to deal with color images or may focus on implementing fast numerical algorithms for SVD evaluation [5].

### References:

- [1] | Kolda, T. G., Bader, B. W. (2009). Tensor decompositions and applications. SIAM rev., 51(3), 455-500.
- [2] | De Lathauwer, L., De Moor, B., Vandewalle, J. (2000). A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. Appl., 21 (4), 1253-1278.
- [3] | Rajwade, A., Rangarajan, A., Banerjee, A. (2012). Image denoising using the higher order singular value decomposition. IEEE PAMI, 35 (4), 849-862.
- [4] | <http://sipi.usc.edu/database/>
- [5] | Halko, N., Martinsson, P. G., Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM rev., 53 (2), 217-288.

Professeur référent : Rodrigo Cabral Farias  
Mail : cabral@i3s.unice.fr

## COVID-19 Incidence Rates Visualization and Interpolation

Santé Publique France makes available to the public daily processed data on the incidence rates of COVID-19 contaminations at different geographic scales: groups of neighboring cities, cities and districts [1]. The main objective of this project is to generate an animation based on these data showing the spatio-temporal evolution of the epidemic at a given geographic scale districts for example during the first month of the autumn's lockdown (from 27/10/2020 to 26/11/2020). The animation will correspond to the daily sequence of choropleth maps of incidence rates, for an example of what should be displayed in one element of this sequence using district data.

All the processing of the data and generation of the choropleth maps and of the animation should be coded in python. The students will have to search for and learn how to use python libraries specially designed for visualizing geographic data.

A second task in this project is to fit a spatial interpolation model on the data to test whether at a given day it is possible to predict the incidence rate at a district based on the incidences from other districts. The model that will be considered will be a spatial Gaussian process [3]. After fitting the model and testing its prediction accuracy through cross validation, a choropleth map containing the obtained interpolation and eventually a map containing the associated uncertainties should be produced.

### References:

- [1] | [https://www.data.gouv.fr/fr/datasets/donnees-de-laboratoires-infra-departementales-durant-lepidemie-covid-19/#\\_](https://www.data.gouv.fr/fr/datasets/donnees-de-laboratoires-infra-departementales-durant-lepidemie-covid-19/#_)
- [2] | <https://www.data.gouv.fr/fr/reuses/repartition-infra-communale-du-covid-19-en-france-metropolitaine/>
- [3] | Williams, C. K., Rasmussen, C. E. (2006). Gaussian processes for machine learning (Vol. 2, No. 3, p. 4). Cambridge, MA: MIT press.

## Visualisation et analyse de modèles dans les réseaux de régulation génétique

Les réseaux de régulation génétique décrivent les interactions entre gènes, protéines, ? (voir [https://en.wikipedia.org/wiki/Gene\\_regulatory\\_network](https://en.wikipedia.org/wiki/Gene_regulatory_network)). Dans la plupart des cas, le comportement statique de tels systèmes est connu (i.e. on connaît les influences individuelles entre les gènes), mais des questions sur la dynamique du système, c'est à dire son comportement global au fil du temps, restent souvent ouvertes. L'équipe Bio-informatique Formelle du laboratoire I3S, travaille avec des chercheurs en biologie, et propose des approches formelles grâce auxquelles les bio-informaticiens aident les biologistes à trancher parmi plusieurs interprétations de leurs résultats expérimentaux (la biologie restant avant tout une science expérimentale). Ces approches sont mises en oeuvre dans l'outil TotemBioNet <https://gitlab.com/totembionet/totembionet>.

L'équipe étudie actuellement les impacts du cancer du pancréas sur la régulation du métabolisme énergétique. Dans ce cadre, des matrices de validation et de prédiction permettent de spécifier et de vérifier des comportements dynamiques du réseau, comme par exemple le fait qu'un marqueur de la mort cellulaire passe un certain seuil. Les sorties de l'outil sont des fichiers excel, contenant les modèles (i.e. les paramètres du réseau) pour lesquels les propriétés sont vérifiées. Pour l'instant l'interprétation de ces résultats se fait manuellement. Récemment, l'utilisation d'une bibliothèque de MDD (Multi Valued Binary Diagrams) a permis d'avoir une représentation synthétique des résultats, où les paramètres qui n'interviennent pas dans un modèle sont abstraits par un « - » dans le fichier excel.

Ce projet a deux objectifs :

- Objectif technique : il s'agit de proposer une visualisation graphique de l'ensemble des paramètres sous la forme d'un arbre de décision. Outre le défi technique de la réalisation de l'IHM, il s'agira de dialoguer avec les bio-informaticiens pour déterminer avec eux l'ordre dans lequel les variables du modèle doivent apparaître dans l'arbre,
- Objectif théorique : il s'agit d'effectuer un état de l'art afin de proposer des méthodes d'analyse mathématiques pour comparer les modèles ou les regrouper en classes. Un point de départ pourrait s'inspirer d'un calcul de similarités sur des données booléennes : <https://journocode.com/2016/03/similarity-and-distance-in-data-part-2/>

Le langage utilisé sera de préférence java ou python.

**Bibliographie** : Quelques éléments sur l'épigénétique et les gènes : <https://www.youtube.com/watch?v=igf6MLmkBcw>, [https://fr.wikipedia.org/wiki/Acide\\_d%C3%A9soxyribonucl%C3%A9ique](https://fr.wikipedia.org/wiki/Acide_d%C3%A9soxyribonucl%C3%A9ique) Présentation générale de l'approche : <https://hal.archives-ouvertes.fr/hal-03027788/document>

Professeur référent : Didier Auroux

Mail : Didier.Auroux@unice.fr

## Observateurs pour l'assimilation de données

L'assimilation de données est l'ensemble des techniques qui permettent de combiner un modèle et des observations (ou données). D'un côté, le modèle, qui est généralement représenté sous forme d'équations mathématiques : c'est la phase de modélisation, d'un phénomène physique, biologique, chimique, etc, qui consiste à représenter ce phénomène à l'aide d'équations mathématiques. Et de l'autre, les données, représentant une source d'information expérimentale ou observationnelle. Et le but de combiner modèle et données est par exemple de reconstituer l'état de l'écoulement d'un fluide géophysique - par exemple un océan, ou l'atmosphère. Une difficulté réside dans la taille du problème, avec plusieurs centaines de millions de données, et un état à reconstruire de dimension plusieurs (dizaines de) milliards.

Une des premières méthodes d'assimilation de données à avoir été utilisée en géophysique est le nudging, aussi connu sous le nom d'observateur de Luenberger. Cet observateur consiste à ajouter directement dans les équations du modèle, un terme de rappel pour tirer la solution vers les observations. Il s'agit d'un observateur asymptotique, au sens où sous certaines conditions, la solution converge asymptotiquement en temps vers la trajectoire réelle que l'on cherche à reconstruire.

De tels observateurs ont été construits et utilisés pour des problèmes d'assimilation de données en géophysique externe (prévisions météo, océanographie, ...). L'idée de ce projet consiste ici à construire et tester numériquement des observateurs pour une équation des ondes dans un contexte géophysique interne, du type détection dans le sous-sol.

Professeur référent :Didier Auroux

Mail : Didier.Auroux@unice.fr

## Couverture d'un actif risqué dans un portefeuille

Un client nous confie une somme d'argent qui sera placée sur un actif sans risque, avec un taux d'intérêt connu, et sur un actif risqué, dont l'évolution à chaque instant peut être à la hausse ou à la baisse. On se placera dans un cadre simplifié avec le modèle de Cox-Ross-Rubinstein, où la variation de l'actif risqué est dicté par une loi binomiale: avec une probabilité  $p$ , le prix est multiplié par un coefficient plus grand que 1, et avec une probabilité  $1-p$ , le prix est multiplié par un coefficient plus petit que 1.

La problématique consiste à s'assurer que le portefeuille aura toujours une valeur suffisante pour satisfaire le client. Etant donné que l'actif sans risque produira de l'argent en permanence, mais que l'actif risqué peut systématiquement en perdre, on se pose la question de la répartition de l'argent du client à effectuer au départ entre ces 2 actifs, afin de garantir une certaine couverture du portefeuille à chaque instant.

Il s'agit ici de faire des simulations numériques, par exemple avec une méthode de Monte-Carlo, pour vérifier les résultats connus sur ce modèle. On cherchera également les limites de ce modèle pour couvrir un portefeuille en fonction des paramètres (taux d'intérêt garanti, et coefficients multiplicatifs de l'actif risqué).

Professeur référent : Didier Auroux

Mail : Didier.Auroux@unice.fr

## Modélisation simple du trafic routier

On s'intéresse à la position et vitesse de plusieurs véhicules se suivant sur une même file. On suppose que les véhicules accélèrent lorsqu'ils ont le champ libre, ne font rien lorsqu'ils sont proches du véhicule de devant, et freinent d'autant plus fort qu'ils en sont très proches. On peut alors placer une dizaine de véhicules les uns derrière les autres, le premier véhicule ayant une vitesse constante, et observer le comportement des véhicules qui suivent.

Le but du projet est de modéliser le comportement à l'aide d'équations différentielles ordinaires couplées, de mettre en oeuvre cette modélisation simpliste, et d'étudier ce qui se passe en fonction des différents paramètres: distances critiques, accélérations/freinages maximum, nombre de véhicules, ...

Une autre option consiste à étudier un modèle continu, à base d'EDP, avec pour inconnues la vitesse des véhicules et leur densité.

On pourra notamment mettre en évidence le phénomène classique d'oscillations dans la distance entre véhicules se suivant, voire l'accident si les véhicules ont des caractéristiques de freinage/accélération trop différents, ou des caractéristiques classiques de bouchons "reculant" sur la route, ...