

Interpolaspline

Rapport du stage applicatif

L3 MIN

Décembre 2019

CORBILLE Clément, DOUMBOUYA Mohamed, EL BOUCHOUARI Zakaria,
HEDDIA Bilel, PIASENTIN Béryll, RODET Amélys

Table des matières

1	Organisation de la semaine	2
2	Contexte et objectifs	3
3	Définition des tâches	3
3.1	Création des tâches	3
3.2	Contenu des tâches	4
4	Cahier des charges	5
5	Planification des tâches	6
5.1	Organisation des tâches	6
5.2	Répartition des tâches	6

Introduction

Dans le cadre de notre troisième année de licence mathématiques-informatique à l'Université Grenoble Alpes, nous devons effectuer un stage applicatif à l'UFR IM²AG. Ce stage, qui consiste en un projet, est une étape indispensable pour mettre en pratique les connaissances acquises durant cette formation et pour acquérir de nouvelles compétences de recherche. Les projets ont lieu durant quatre semaines, réparties en deux périodes de respectivement une puis trois semaines. Des groupes de six étudiants ont été construits en fonction des affinités de chacun pour les différents sujets proposés.

Le sujet de notre projet est l'interpolation avec données aberrantes. Il nous donne l'opportunité d'approfondir nos connaissances en Mathématiques appliquées, plus précisément d'approfondir l'enseignement d'algèbre linéaire pour le graphique et la CAO (Conception Assistée par Ordinateur) que nous avons reçue, ainsi que la programmation en langage python. De plus, ce stage va nous permettre d'acquérir des compétences en travail de groupe.

1 Organisation de la semaine

L'objectif de cette semaine de stage était de mettre en place l'organisation nécessaire à la réalisation du projet, qui aura lieu en avril.

Nous avons commencé par désigner un chef de projet : Béril a été choisie à cet effet. La chef de projet veille sur le bon déroulement de chaque tâche en plus de sa participation aux différentes tâches. Nous avons ensuite défini un nom pour le projet, en rapport avec le contenu mais également percutant. "Interpolaspline" a été le nom retenu, en tant que fusion du mot "interpolation" et du mot "spline", deux notions principales du sujet.

Durant cette semaine d'organisation, chaque membre du groupe a été responsable d'une ou de plusieurs tâches. Être responsable d'une tâche signifie s'assurer que celle-ci est correctement réalisée et terminée dans les délais fixés, mais pas obligatoirement qu'elle doit être réalisée seul.

Les premiers jours ont été mis à profit pour comprendre le sujet et trouver ce sur quoi nous allions travailler pendant trois semaines. Il a fallu ensuite établir un plan de travail. Cette réflexion nous a permis de construire un diagramme de Gantt et un graphe de dépendances des tâches, mais aussi de répartir les tâches entre les différents membres du groupe. Chacun sait donc de quelle tâche il est responsable et peut se projeter dans la semaine de réalisation.

Une fois tous ces points mis en place, nous avons établi le cahier de charges. Nous y avons décrit le problème, nos objectifs, le cadre du projet. La description des fonctionnalités livrées à la fin du projet y est également détaillée, ainsi que l'organisation temporelle du projet.

Tous les membres des groupes ont participé à la rédaction du rapport basé sur cette semaine de travail, incluant le cahier des charges, mais aussi à la préparation de la soutenance. Pour terminer cette semaine d'organisation, chacun a pris le temps de bien comprendre ce qu'il devrait faire durant les trois semaines de réalisation.

2 Contexte et objectifs

REDIGER DU BLABLA SUR QQQQCCP

3 Définition des tâches

3.1 Création des tâches

Dans le but d'interpoler des données en utilisant des splines de lissage tenant compte des valeurs aberrantes, nous nous devons de savoir comment aborder le sujet. En discutant entre membres du groupe, nous avons convenu qu'il fallait dans un premier temps approfondir les notions générales sur les splines naturelles et de lissage en fonction de la répartition des points et trouver un moyen d'automatiser l'estimation du paramètre de lissage. Puis nous avons prévu d'étudier des méthodes d'identification des valeurs aberrantes que l'on souhaite à terme implémenter en deux dimensions, mais aussi des méthodes permettant d'interpoler des splines sans commencer par détecter les valeurs aberrantes.

3.2 Contenu des tâches

Les tâches sont les suivantes :

- Approfondissement des splines naturelles
- Approfondissement des splines de lissage uniformes
- Approfondissement des splines de lissages non-uniformes
- Estimation automatique du paramètre de lissage
- Identification des points aberrants
- Etude des splines de lissage avec points aberrants
- Etude de l'algorithme RanSac
- Redéfinition d'autres fonctions d'erreur à minimiser
- Finalisation du projet

Voici quelques explications du contenu scientifique, caractérisant les tâches listées ci-dessus.

Splines naturelles

L'interpolation est une opération mathématique consistant à déterminer une fonction passant par un nombre fini de points donnés.

Soient (x_i, y_i) les $n + 1$ données, $i \in \llbracket 0; n \rrbracket$. Une spline est une fonction définie par morceaux sur chaque intervalle $[x_i, x_{i+1}]$, $i \in \llbracket 0; n - 1 \rrbracket$. Le principe d'une spline est de combiner plusieurs morceaux de fonctions (en général on utilise des polynômes de petit degré) pour former une fonction d'une classe définie.

Les splines cubiques C^2 sont les splines définies par des polynômes cubiques dont la dérivée seconde est continue. On appelle spline naturelle la spline cubique C^2 dont les dérivées secondes aux extrémités de l'intervalle de définition sont nulles. La spline naturelle existe toujours pour des données distinctes, et est unique. Le principe de la spline cubique naturelle permet d'obtenir un système tridiagonal simple qui peut être résolu facilement, afin d'obtenir les coefficients des polynômes.

Splines de lissage

Paramètre de lissage

Identification des points aberrants

Splines de lissage avec points aberrants

RanSac

Minimisation de nouvelles quantités

Paramétrique

L'objectif de l'interpolation paramétrique est de trouver un polynôme paramétrique qui interpole les données. Pour ce faire, deux méthodes sont possibles :

- L'interpolation séparée des n données (t_i, x_i) et des n données (t_i, y_i) , $i \in \llbracket 0; n \rrbracket$, avec t un paramètre bien choisi. En associant les deux splines obtenues grâce au paramètre t , on obtient une spline paramétrique interpolant les données de départ.
 - L'interpolation e
- Soit on interpole les données directement, en cherchant un polynôme paramétrique qui passe par les points de coordonnées (X_i, Y_i) pour $i = 1, \dots, N$. - Soit en interpolant deux polynômes Q et R par les

données (t_i, X_i) et (t_i, Y_i) pour $i = 1, \dots, N$, avec t_i est un paramètre bien choisi. En regroupant les résultats, on obtient un polynôme paramétrique P qui interpole les données (X_i, Y_i) . $P(t) = (Q(t), R(t))$ Avec $P(t_i) = (Q(t_i) = X_i, R(t_i) = Y_i)$ pour $i = 1, \dots, N$.

Finalisation du projet

L'interpolation paramétrique :

Spline de lissage : Les splines de lissage sont des estimations de fonction, obtenues à partir d'un ensemble d'observations bruyantes Y_i de la cible $f(X_i)$, afin de trouver la spline minimale qui suit le nuage des points (X_i, Y_i) mais qui passe pas forcément par tous les points, cette méthode vise à éviter l'oscillation de la courbe de la spline quand n est très grand - Le paramètre de lissage : La quantité à minimiser est composée de deux parties, la première assure le passage de la spline par nos données et l'autre assure le lissage de la courbe. Donc plus qu'on augmente la valeur du paramètre de lissage plus qu'on donne un poids à la partie de lissage, ce qui nous donne par suite une spline continue et lisse.

Données aberrantes : En statistique, une donnée aberrante est une valeur ou une observation qui est « distante » des autres observations effectuées sur le même phénomène, c'est-à-dire que c'est une valeur qui s'écarte fortement des valeurs des autres observations anormalement faible ou élevée Sur un graphe on peut les visualisées comme des points qui n'appartiennent pas au nuage des points totales - Détermination des points aberrants : Il existe plusieurs méthodes variées pour la détection des données aberrantes ; on a prévenu plusieurs méthodes de recherche on implémente ceux qui sont plus intéressante en fonction du temps . Cependant il y a des méthodes qui suppriment les points aberrants sans les détecter comme l'algorithme RANSAC qui s'agit d'un algorithme non-déterministe dans le sens où il produit un résultat correct avec une certaine probabilité seulement .

Conclusion : L'objectif de ce projet c'est de développer et adapter des méthodes qui interpellent des points donnés en utilisant des splines de lissages qui sont basé sur la minimisation des quantités dont les moindres carrées et d'autre qu'on le précisera par la suite, en traitant les cas des points aberrants et en trouvant le meilleurs paramètre de lissage .

4 Cahier des charges

Le cahier des charges est l'annexe n° 4. Nous y avons détaillé le projet, du sujet à l'archive rendue, en passant par sa réalisation. Ce cahier des charges représente notre compréhension globale du sujet ainsi que les principales idées d'organisation et de réalisation.

Ce document est censé être le lien entre notre équipe et le client, mais il est ici le lien entre notre équipe et l'équipe d'enseignants responsables du stage applicatif. Nous allons l'utiliser durant la seconde fin du projet. Notre objectif est de réaliser tous les objectifs qui y sont définis, dans les délais impartis.

5 Planification des tâches

5.1 Organisation des tâches

5.2 Répartition des tâches

Conclusion

Conclusion : Lors de cette semaine travail nous avons remarqué qu'au final, nous avons eu une bonne cohésion de groupe. Etant donné qu'au début du projet nous avons tous des idées différentes, nous avons toutefois réussi à nous mettre d'accord sur une trame de projet. Cette cohésion n'était pourtant pas évidente lors de la formation de notre groupe puisque les groupes sont formés en fonction des sujets. Afin de suivre cette trame, nous avons donc dû organiser. C'est-à-dire que nous avons partagé le travail en groupe en sorte que chacun ait au moins une tâche, afin de gagner du temps. Nous nous sommes même découvert une âme d'acteur pour chacun d'entre nous, c'est pourquoi selon les caractères et personnalités de chacun nous avons attribué les rôles afin que cela soit le plus cohérent possible.

Annexes