EISTI - DÉPARTEMENT MATHÉMATIQUES

ANALYSE NUMÉRIQUE II - TP N° 2

15 février 2009

INTRODUCTION

Une distinction doit être faite entre approximation, interpolation et extrapolation d'une fonction. Le cadre commun à ces trois méthodes est un ensemble de points dans un intervalle pour lesquels on connaît les valeurs de la fonction.

- **Approximation** : On admet ici que les valeurs de points connus puissent comporter des imprécisions. La fonction d'approximation ne prend pas exactement les valeurs des points mais tend à minimiser l'écart avec ceux-ci. La méthode des moindres carrés (cf. §5.6, page 118-119) est un exemple d'approximation.
- **Interpolation** : Ici la fonction passe par l'ensemble de points connus. Autrement dit, l'interpolation consiste à trouver une fonction qui correspond aux valeurs de points connus et qui est capable de donner des valeurs pour des points situés entre deux points connus consécutifs.
- **Extrapolation** : Soit une fonction passant par un ensemble de points dans un intervalle. On s'intéresse ici aux valeurs que peut prendre la fonction d'extrapolation en dehors de l'intervalle des points connus. Une extension intéressante de l'extrapolation, que nous allons considérer dans ce TP, est la prédiction. À l'aide des points connus, on établit un modèle mathématique qui simule le comportement de la fonction inconnue. En utilisant ce modèle on peut établir des prédictions sur la valeur de fonction à l'instant $m+k$ en se fondant sur les valeurs de la fonction jusqu'à l'instant m. Comme méthode de prévision on peut utiliser la méthode des moindres carrés.

L'objectif de ce TP est de manipuler des outils de développement pour construire des fonctions d'interpolation et d'extrapolation à partir d'échantillons numériques concernant les tâches du soleil (sunspot numbers).

TRAVAIL À FAIRE

Vous utiliserez le fichier `soleil1900-83.dat` qui contient le nombre de tâches solaires entre 1900 et 1983, mois par mois.

(1) **Interpolation par splines cubiques**

Nous avons vu, lors du TD précédent, que l'interpolation de Lagrange est de piètre qualité dès que le nombre de points connus est > 5. Une méthode plus intéressante et de bien meilleure qualité est l'interpolation par des fonctions splines. Ces fonctions sont des ensembles de polynômes de faible degré qui définissent localement la fonction étudiée. Un polynôme est défini pour chaque intervalle entre les points de l'ensemble des données. Quand le nombre de points augmente, le nombre de polynômes augmente mais pas leur degré et ainsi on n'observe pas de phénomène de Runge.

Quand les polynômes d'une fonction spline sont d'ordre 1, le système se ramène à une interpolation linéaire. On utilise couramment des ensembles de polynômes d'ordre 3 appélés splines cubiques. Pour les détails mathématiques, il faut se réporter au poly du cours, §7.3, page 154 et suivantes.

(a) Mise en øeuvre de l'interpolation par splines cubiques en Scilab.

i. Scilab dispose d'une fonction `splin` qui calcule les splines cubiques selon un ensemble de points connus. Supposons, pour fixer les idées, que nous connaisons les valeurs de la fonction à n points $\mathbf{x} = [x_1, \ldots, x_n]^\top$. Notons ces valeurs $\mathbf{y} = [y_1, \ldots, y_n]^\top$. L'appel à la fonction spline se fait par intermédiaire de la commande

```
ds = splin x, y]
```

qui calcule la fonction spline s telle que $s(x_i) = y_i; i = 1, \ldots, n$. Le résultat obtenu `ds` est la dérivée de la fonction s aux points x_i.

ii. Ensuite nous allons évaluer la fonction spline dans différents points entre deux points connus et ceci pour tout couple de points connus onsécutifs. Par exemple on prendra 10 points équidistants entre deux points x_i et x_{i+1}. On construit ainsi le vecteur $\mathbf{xx} = \left[x_{1,1}, \ldots, x_{1,10}, x_{2,1}, \ldots, x_{n,10}\right]^\top$. L'interpolation se fera à l'aide de la fonction `interp` qui calcule la valeur de la fonction spline cubique pour les points du vecteur **xx**, par la commande

```
yy = interp xx,x,y,d]
```

(b) Procéder comme indiqué ci-dessus pour faire une interpolation par spline cubique des tâches solaires.

(c) Tracer la courbe des tâches solaires et de l'interpolation.

(d) Commenter les résultas obtenus.

(2) **Prévision par moindres carrés**

Le problème qu'on se pose est le suivant : connaissant le nombre de tâches solaires jusqu'au mois n, prévoir ce nombre pour le mois prochain.Pour ce faire on doit avoir un modèle mathématique qui exprime le nombre de tâches à un mois donné en fonction des tâches de certains mois précédents. En formalisant on peut dire que nous connaissons ces nombres pour N mois consécutifs (dans notre cas $N = 1008$), que l'on note par $x(k); k = 1, \ldots, N$ et nous voulons prévoir le nombre de tâches solaires au mois $m+1$, avec $m+1 \leq N$. Nous allons utiliser la méthode des moindres carrés. Supposons que le modèle que nous voulons établir est le suivant :

$$\widehat{x}(m+1) = a_1 x(m) + \cdots + a_p x(m-p+1)$$

où $\widehat{x}(m+1)$ est la prévision à l'instant m du nombre de tâches solaires à l'instant $m+1$. Il s'agit d'un modèle linéaire d'ordre p. On pourra calculer $\widehat{x}(m+1)$ si on connaît les valeurs des coefficients $a_1, \ldots, a_p$. Pour effectuer leur calcul on utilisera la méthode des moindres carrés dont on rappelle la forme : trouver le vecteur $\mathbf{a} \in \mathbb{R}^L$ qui approche au mieux la solution du système sur-déterminé

$$\mathbf{Xa} = \mathbf{y}; \quad \mathbf{X} \in \mathbb{R}^{K \times L}, \quad K > L$$

La solution est donnée par la pseudoïnverse de $\mathbf{X}$, selon la formule

$$\mathbf{a} = \mathbf{X}^{+}\mathbf{y} = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\mathbf{y}$$

(a) Mise en øeuvre de la méthode des moindres carrés pour la prévision.

i. Formation de la matrice des données $\mathbf{X}$ et du vecteur $\mathbf{y}$:

$$\mathbf{X} = \begin{bmatrix} x(1) & x(2) & \cdots & x(p-1) \\ x(2) & x(3) & \cdots & x(p) \\ \vdots & \vdots & \ddots & \vdots \\ x(m-p+1) & x(m-p+2) & \cdots & x(m-1) \end{bmatrix} \quad ; \quad \mathbf{y} = \begin{bmatrix} x(p) \\ x(p+1) \\ \vdots \\ x(m) \end{bmatrix}$$

ii. Calcul, par moindres carrés, du vecteur $\mathbf{a}$

iii. Calcul de la prévision $\widehat{x}(m+1)$ selon la formule $\widehat{x}(m+1) = a_1 x(m) + \cdots + a_p x(m-p+1)$.

(b) Pour faire des prévisions, on établira le modèle pour $m = 900$ mois. Ensuite on calculera la prévision pour $m+1 = 901$ qu'on comparera avec la vraie valeur $x(901)$. Ensuite on continuera en établissant le modèle pour $m = 901, 902, \ldots, 1007$ et on calculera les prévisions pour $m = 902, 903, \ldots, 1008$.

(c) On établira un critère pour évaluer la qualité de la prévision en utilisant la totalité des prévisions.

(d) En utilisant ce critère on cherchera à établir la meilleure valeur pour l'ordre p du modèle. On commencera avec $p = 2$ et on augmentera progessivement en essayant d'améliorer le critère.

(e) Tracer la courbe des tâches solaires et des prévisions pour le meileur ensemble des prédictions.

(f) Commenter les résultas obtenus.

MODALITÉS PRATIQUES

Vous devez rediger un rapport détaillé dans lequel vous expliquerez et justifierez votre travail et de vos choix techniques. En effet, le sujet est volontairement flou quant aux critères à utiliser et vous devez donc justifier et commenter votre travail.

En particulier vous devez présenter et commenter les résultats obtenus.

De plus il faut fournir la liste de vos références.

Vous fournirez aussi les sources de vos programmes, prêts à être exécutés en lançant le programme principal.

Nous vous rappelons que tout plagiat d'un texte existant ou d'un autre rapport vous expose à avoir une note de zéro. En effet c'est votre capacité à vous exprimer sur un sujet scientifique qui nous intéresse, pas vos performances de copieur/colleur.

Une attention particulière sera accordée au soin et à la présentation du rapport ainsi qu'à la qualité et la lisibilité des programmes.

La totalité de vos fichiers sera mis dans un fichier compacter (zip, rar ou gzip)et envoyer en utilisant le bouton correspondant du TP 2 sur le site http ://sifoci.eisti.fr.

Date de remise du rapport : Lundi 9 mars à 23h59.