Abdelkader BENHARI

# ALGORITHMES D'AJUSTEMENT DE COURBE ET DE LISSAGE SOUS DES HYPOTHESES DE CONVEXITE

## ALGORITHMS OF CURVE ADJUSTMENT AND SMOOTHING UNDER CONVEXITY ASSUMPTIONS

Ce livre a été publié sur www.bookelis.com
ISBN : 979-10-227-0371-0

Résumé

Ce travail a été consacré à l'ajustement et au lissage de courbes, nous avons été amenés à fournir des méthodes algorithmiques et à les tester à partir d'une part, d'exemples concrets : trafic sur autoroute, d'autre part, d'exemples simulés : estimation d'une densité de probabilité par la méthode du noyau, régressogramme.

Les résultats ont été satisfaisants, notamment sur le lissage de l'estimation de densité de probabilité et de fonction de régression, ainsi que pour l'ajustement du nuage de points (modèle paramétrique). En d'autres termes, cette méthode peut, à la fois s'appliquer à de nombreux modèles paramétriques, et permettre le lissage d'autres estimateurs que ceux traités dans cette étude.

Abstract

This work is devoted to the adjustment and smoothing of curve, we were asked to provide algorithmic methods and test them from one hand, a concrete example: traffic on the motorway, on the other hand, to simulated example: estimation of probability density by the kernel method, smooth regression.

The results were satisfactory, including the smoothing of the estimated probability density and regression function and for adjustment of the point cloud (parametric model).

In other words, this method can both be applied to many parametric models, and to allow smoothing of other estimators than those treated in this study.

In fact, this work leaves open the question of adjustment of a cloud of points, although this adjustment can be improved by adding additional constraints on the slopes of the segments forming the polyhedral surface adjustment.

keywords

Algorithm, mathematic, adjustment and smoothing, estimate statistics, nonparametric estimation, kernel method, regression, probability density function regression, computer statistical, estimator,

Sommaire

## Introduction

Le problème d'ajustement de courbes, est très proche des questions liées au traitement de données expérimentales.

Si on a un nuage de points $(x_i, y_i)$, ( i=1,2,…,n ), issus d'une expérience qui a pour but d'étudier la relation entre une grandeur physique y ,et une grandeur physique x ; sous des considérations théoriques ou physiques, on suppose que x et y sont liés par la relation y = f(x, a, b, c, …). C'est cette relation qu'il y a lieu de trouver, à partir des données $(x_i, y_i)$, (i=1,2,…,n), c'est-à-dire ajuster au nuage de points une fonction y = f(x, a, b, c, …) qui dépend des paramètres a, b, c, …

La question qui se pose est :

Comment, à partir du nuage de points, trouver f(x, a, b, c, …) avec la meilleure approximation possible ?

Pour résoudre ce problème, nous utilisons, dans ce travail, des méthodes de lissage par des lignes polygonales convexes ou concaves.

Comme dans tout problème d'ajustement, il y a lieu d'établir ce que peut être la meilleure solution ; nous considérons donc :

1) La meilleure : la solution assurant d'une part, le minimum du maximum des valeurs absolues des écarts, d'autre part, le minimum de la valeur absolue des sommes des écarts entre la courbe d'ajustement et le nuage de points.
2) La meilleure : la solution où l'on veut que soit minimale la somme des valeurs absolues des écarts, entre la courbe d'ajustement et le nuage.

La première partie de cette étude, consiste à donner des algorithmes d'ajustement de courbes.

Dans la deuxième partie, nous appliquerons ces algorithmes à un problème paramétrique semblable au problème décrit ci-dessus. Nous proposons également, une manière dont on peut déterminer les paramètres des modèles considérés.

Dans les troisième et quatrième partie, nous emploierons ces méthodes pour lisser des estimateurs de densités de probabilités par la méthode du noyau, et pour lisser le régressogramme comme estimateur d'une fonction de régression.

Enfin, pour tester nos méthodes, nous avons mis au point, plusieurs programmes informatiques pour effectuer le traitement de données sur ordinateur.

Première partie
Ajustement d'une courbe à un nuage de points en utilisant
des hypothèses de convexité

Les techniques de lissage des courbes ou des nuages de points ont fait, depuis quelques années, l'objet de nombreux travaux et ont été très étudiées en littérature.
Nombreuses sont les représentations graphiques de fonction et d'estimateurs obtenus empiriquement dans différents domaines et particulièrement en statistiques, ces représentations ont besoin d'être lissées et peuvent l'être.
L'intérêt des procédures de lissage s'est accru avec le développement des moyens de calcul et spécialement les possibilités de représentations graphiques (traceurs, terminaux graphiques, …) qui simplifient les lourdes tâches de recherche des critères de comparaison et qui « font parler les données d'elles mêmes ».

a) Nuage de points ayant une forme convexe
Le but ici est de présenter quelques façons d'utiliser des hypothèses de convexité ou de concavité pour ajuster un nuage de points.
Supposons que l'on ait un nuage de points $(X_j, Y_j)_{1\le j\le n}$ dont la forme générale est convexe. Dans la recherche d'une ligne polygonale convexe dont les sommets seront les points de coordonnées $(x_i, z_i)_{0\le i\le k}$, k un entier dépendant de n, qui ajuste le nuage de points, l'hypothèse de convexité intervient de la façon suivante :

$$(1)\quad \frac{z_{i+1}-z_i}{x_{i+1}-x_i} \ge \frac{z_i-z_{i-1}}{x_i-x_{i-1}} \quad pour\ \ i \in \{1,2,...,k-1\}$$

b) Nuage de points concave.
Dans le cas d'un nuage de points concave, l'hypothèse de concavité de la ligne polygonale ajustée au nuage est :

$$(2)\quad \frac{z_{i+1}-z_i}{x_{i+1}-x_i} \le \frac{z_i-z_{i-1}}{x_i-x_{i-1}} \quad pour\ \ i \in \{1,2,...,k-1\}$$

c) Nuage de points ayant une forme tourmentée
On peut avoir un nuage de points $(X_j, Y_j)_{1\le j\le n}$ ayant une forme plus tourmentée, c'est-à-dire convexe en certaines parties concave en d'autres et même rectiligne ailleurs. Les parties rectilignes ne posent pas de problème, nous supposons qu'il n y a pas de ces parties dans le nuage de points. Dans ce cas, on divise le nuage de points

en plusieurs parties, de manière à ce que la forme de chacune d'elle soit convexe ou concave et l'on fait intervenir pour chacune d'elle l'hypothèse (1) ou l'hypothèse (2).

d) Réduction du nuage de points

Quand la taille n d'un nuage de points est grande, plutôt que d'utiliser tous les points du nuage, il est plus simple de le réduire en un nombre beaucoup plus faible de points qui donne une reconstitution presque aussi satisfaisante que le nuage de points initial ; deux méthodes parmi d'autres sont possibles.

Pour un nuage de points de coordonnées $(X_j, Y_j)_{1\le j\le n}$ on trouve le plus petit des $(X_j)_{1\le j\le n}$, $x_{\min}$, et le plus grand $x_{\max}$ ; nous divisons l'intervalle $\left[x_{\min}, x_{\max}\right]$ en un nombre k d'intervalles, $\Delta_{n,r}$ $r \in \{1,2,...,k\}$, de même longueur.

On note :

$x^0, x^1, x^2, ..., x^k,$ $avec$ $x^0 = x_{\min}$ , $x^k = x_{\max}$ les bornes des intervalles $\Delta_{n,r}$ $r \in \{1,2,...,k\}$

i) Réduction par la méthode des moyennes

On obtient un nuage de points réduit, en remplaçant $y_j$ par la moyenne des ordonnées des points du nuage initial tombant dans l'intervalle $\left[x^i, x^{i+1}\right]$ $avec$ $i = 0,1,2,...,k-1$

ii) Méthode utilisant les moindres carrés

Par la méthode des moindres carrés, on ajuste un segment de droite aux points tombant dans l'intervalle $\left[x^i, x^{i+1}\right]$ $avec$ $i = 0,1,2,...,k-1$

Pour i=0 (respectivement i=k), on prend le point d'abscisse $x^0$ (respectivement $x^k$) du segment correspondant à l'intervalle $\left[x^0, x^1\right]$ (respectivement $\left[x^{k-1}, x^k\right]$).

Pour $1 \le i \le k-1$, on prend le milieu des extrémités des deux segments correspondant à $\left[x^{i-1}, x^i\right]$ $et$ $\left[x^i, x^{i+1}\right]$.

e) Critères d'optimisation

Nous employons, dans cette étude, divers critères d'optimisation pour ajuster une ligne polygonale à un nuage de points et lisser un estimateur de la densité de probabilité ainsi que le régressogramme comme estimateur de la fonction de régression. Les solutions que nous proposons à ce problème emploient les critères d'optimisation suivants, sous une formulation de programme linéaire.

i) Minimisation du maximum de la valeur absolue des écarts.
Ceci constitue la première étape d'une première méthode alors que la minimisation de la somme des écarts en valeur absolue constitue la seconde étape.

ii) Minimisation des sommes des valeurs absolues des écarts.

f) <u>Lissage d'une ligne polygonale.</u>

i) A partir d'un nuage de points $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ (obtenu réellement ou par simulation), on détermine un nuage de (k+1) points par l'une des méthodes de réduction. Nous faisons ensuite un lissage des points $(z_i)_{0 \le i \le k}$ de manière à obtenir

$(\overset{\frown}{z_0}, \overset{\frown}{z_1}, \overset{\frown}{z_2}, \ldots, \overset{\frown}{z_k})$ , ordonnées des sommets d'une ligne polygonale C joignant $(x^0, \overset{\frown}{z_0}), (x^1, \overset{\frown}{z_1}), \ldots, (x^k, \overset{\frown}{z_k})$ .

On leur impose l'une des conditions (1) ou (2) suivant la forme du nuage en appelant $\Psi$ la fonction ayant pour graphe C :

(1) $\Psi$ convexe (si l'on sait que le nuage est convexe)

(2) $\Psi$ concave (si le nuage est concave)

Posons

$$z = \max_{0 \le i \le k} \left| \overset{\frown}{z_i} - z_i \right|$$

La minimisation de ( z ) sous la condition (1) ou (2) est un programme linéaire ; en en effet si $\delta = z$ on a :

$$(2) \quad \left[ \begin{array}{ll} \delta = z \ \min & \\ \left| \overset{\frown}{z_i} - z_i \right| \le \delta & avec \quad i = 0, 1, 2, \ldots, k \\ \overset{\frown}{z_{i+1}} - \overset{\frown}{z_i} \le \overset{\frown}{z_i} - \overset{\frown}{z_{i-1}} & avec \quad i = 1, 2, \ldots, k-1 \\ \delta \ge 0 & \end{array} \right.$$

Ou

$$
\begin{cases}
\delta = z \ \min & \\
\left|\overset{\frown}{z_i} - z_i\right| \leq \delta & avec \quad i = 0,1,2,\ldots,k \\
(1) \quad \overset{\frown}{z_{i+1}} - \overset{\frown}{z_i} \geq \overset{\frown}{z_i} - \overset{\frown}{z_{i-1}} & avec \quad i = 1,2,\ldots,k-1 \\
\delta \geq 0 &
\end{cases}
$$

Si $\Psi$ a une forme tourmentée et change de sens de courbure au point $x^r$, supposons qu'elle est convexe dans $\left[x_{\min}, x^r\right]$ et concave sur $\left[x^r, x_{\max}\right]$, on aura le programme suivant :

$$
\begin{cases}
\delta = z \ \min & \\
\left|\overset{\frown}{z_i} - z_i\right| \leq \delta & avec \quad i = 0,1,2,\ldots,k \\
(1) \quad \overset{\frown}{z_{i+1}} - \overset{\frown}{z_i} \geq \overset{\frown}{z_i} - \overset{\frown}{z_{i-1}} & avec \quad i = 1,2,\ldots,r-1 \\
(2) \quad \overset{\frown}{z_{i+1}} - \overset{\frown}{z_i} \leq \overset{\frown}{z_i} - \overset{\frown}{z_{i-1}} & avec \quad i = r, r+1, r+2,\ldots,k-1 \\
\delta \geq 0 &
\end{cases}
$$

La résolution de l'un des programmes précédents fournit les $(\overset{\frown}{z_i})$, $i \in \{0,1,2,\ldots,k\}$, donc $\Psi$, ainsi que le minimum de z ; nous appellerons cette valeur $z_0$.

ii) Il est évident que les programmes précédents sont toujours réalisables. Par exemple, dans l'hypothèse où C doit être concave, toute ligne polygonale concave dont les sommets ont pour abscisses $x^0, x^1, \ldots, x^k$ est une solution réalisable.
D'autre part, nous avons constaté sur divers exemples qu'un tel programme peut admettre une infinité de solutions. Il semble que ce soit le cas général.

Nous nous efforcerons alors de trouver, parmi toutes ces solutions, une qui soit bien centrée. Pour cela nous introduisons une seconde étape de la méthode considérée qui a pour objectif de trouver les sommets $(x_i, \overset{\frown\frown}{z_i})_{0 \leq i \leq k}$ d'un polygone $C_2$ tel que :

$$t = \left| \sum_{i=0}^{k} (\widehat{z_i} - z_i) \right| \quad \text{soit minimum}$$

Les conditions imposées aux $(\widehat{z_0}, \widehat{z_1}, ..., \widehat{z_k})$ sont :

$$\begin{cases} \text{les hypothèses } (1) \text{ ou } (2) \\ \text{et} \\ \left| \widehat{z_i} - z_i \right| \leq z_0 \quad pour \ \ i = 0,1,2,...,k \end{cases}$$

Ce qui nous donne le programme linéaire suivant à résoudre :

$$\left[ \begin{array}{l} \lambda = t \ (\min) \\ \left| \widehat{z_i} - z_i \right| \leq z_0 \qquad avec \ \ i = 0,1,2,...,k \\ \left| \sum_{i=0}^{k} (\widehat{z_i} - z_i) \right| \leq \lambda \\ \widehat{z_{i+1}} - \widehat{z_i} \geq \widehat{z_i} - \widehat{z_{i-1}} \quad avec \ \ i = 1,2,...,k-1 \\ \lambda \geq 0 \end{array} \right.$$

Dans le cas d'un échantillon concave ou ayant une forme tourmentée, seule la contrainte apparentée à cette condition change.

Une fois résolu, ce programme nous donne une fonction $\widehat{\Psi}$ pouvant mieux ajuster notre nuage.

<u>Remarque :</u>

L'utilisation du premier critère d'optimisation seul ( $z = \max\limits_{0 \leq i \leq k} \left| \widehat{z_i} - z_i \right|$ ) semble être insuffisante.

iii) Pour l'utilisation du troisième critère d'optimisation nous aurons à déterminer $(x_i, \widehat{z_i})_{0 \leq i \leq k}$ tel que :

$$\alpha = \sum_{i=0}^{k} \left| \widehat{z_i} - z_i \right|$$

Soit minimum sous les conditions de convexité (1) ou (2) ceci est encore un programme linéaire ; soit $(\delta_i)_{0 \le i \le k}$ une suite de réels positifs, la minimisation de $\alpha$ revient à minimiser $\sum_{i=0}^{k} \delta_i$ tel que :

$$\left[\begin{array}{ll} \left| \widehat{z_i} - z_i \right| \le \delta_i & avec \quad i = 0,1,2,...,k \\ \widehat{z_{i+1}} - \widehat{z_i} \ge \widehat{z_i} - \widehat{z_{i-1}} & avec \quad i = 1,2,...,k-1 \\ ou & \\ \widehat{z_{i+1}} - \widehat{z_i} \le \widehat{z_i} - \widehat{z_{i-1}} & avec \quad i = 1,2,...,k-1 \end{array}\right.$$

Ce procédé nous donne une deuxième méthode pour arriver à notre but ( obtenir un polygone ajustant un nuage de points)

Deuxième partie
Application à l'ajustement d'un modèle paramétrique
( Modèle d'écoulement de trafic sur autoroute )

I) <u>Théorie hydrodynamique de l'écoulement du trafic</u>

i) <u>Notions fondamentales</u>

La circulation des véhicules présente certaines analogies avec celle des molécules d'un fluide. C'est pourquoi, la théorie hydrodynamique permet une approche du problème. On note cependant quelques différences :

- Les véhicules ne sont pas élastiques
- Les véhicules cherchent à s'éviter.

Nous utilisons, dans cette partie, les notations suivantes :

$q(x,t)$ : Le débit à l'abscisse x et à l'instant t est le rapport $\frac{\Delta n}{\Delta t}$ où $\Delta n$ est le nombre de véhicules qui passent à l'abscisse x pendant l'intervalle de temps $\Delta t$ (situé autour de t)

$k(x,t)$ : La concentration à l'abscisse x et à l'instant t est le rapport $\frac{\Delta n}{\Delta x}$ où $\Delta n$ est le nombre de véhicules qui se trouvent à l'instant t sur le tronçon de longueur $\Delta x$ .

$h(x,t)$ : L'écart moyen entre deux véhicules à l'instant t et à l'abscisse x, on a :

$$h(x,t) = \frac{1}{q(x,t)}$$

$s(x,t)$ : L'écart spatial moyen entre deux véhicules, à l'instant t et l'abscisse x :

$$s(x,t) = \frac{1}{k(x,t)}$$

Soit $u = \frac{q}{k}$ on montre que u est égal à la vitesse moyenne spatiale et à la moyenne harmonique temporelle.

Par la suite, on appellera vitesse cette quantité : $u(x,t) = \frac{q(x,t)}{k(x,t)}$

ii) Equation de conservation.

Considérons le tronçon $\left[ x,\, x + \Delta x \right]$, à l'instant t, il contient $n(t) = k(x,t)\Delta x$ véhicules.

A l'instant $t + \Delta t$, il contient $n(t + \Delta t)$ véhicules avec :

$n(t + \Delta t) = k(x, t + \Delta t)\Delta t$

Or

$$k(x, t + \Delta t) \simeq k(x,t) + \frac{\partial k}{\partial t}(x,t)\Delta t$$

Donc le gain de véhicules sur le tronçon $\left[x, x + \Delta x\right]$, entre $t$ et $t + \Delta t$ est :

$$n_g = n(t + \Delta t) - n(t) \simeq \frac{\partial k}{\partial t}(x,t)\Delta t \Delta x$$

Soit $n_e$ et $n_s$, le nombre de véhicules entrés entre $t$ et $t + \Delta t$ sur le tronçon $\left[x, x + \Delta x\right]$ et le nombre de véhicules sortis du tronçon $\left[x, x + \Delta x\right]$ entre $t$ et $t + \Delta t$, on a alors :

$n_e = q(x,t)\Delta t$

$n_s = q(x + \Delta x, t)\Delta t$

$$q(x + \Delta x, t)\Delta t \simeq q(x,t)\Delta t + \frac{\partial q}{\partial x}(x,t)\Delta t \Delta x$$

La conservation implique donc que :

$$\frac{\partial k}{\partial t}(x,t)\Delta t \Delta x = n_e - n_s = -\frac{\partial q}{\partial x}(x,t)\Delta t \Delta x$$

D'où

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0$$

iii) Vitesse de propagation de l'onde de propagation

L'hypothèse généralement admise est que :

$$q = q(k)$$

L'équation de conservation devient :

$$\frac{\partial k}{\partial t} + q'(k)\frac{\partial k}{\partial x} = 0$$

Soit

$$c = q'(k) = \frac{dq}{dk}$$

$$q = uk$$

Cela implique que : $c = u + k\frac{du}{dk} < u$ puisque $\frac{du}{dk} < 0$

(1) $\frac{\partial k}{\partial t} + c\frac{\partial k}{\partial x} = 0$ $c < u$

<u>Les solutions de l'équation (1)</u> :

a) Si c ne dépend ni de x, ni de t, la solution est :

$k = f(x - ct)$ , onde de concentration se propageant à la vitesse c.

b) Si c dépend de x et de t , on a :

$$\begin{aligned} dk &= \frac{\partial k}{\partial x}dx + \frac{\partial k}{\partial t}dt \\ &= \frac{\partial k}{\partial x}(dx - cdt) \\ &= 0 \end{aligned}$$

Car k est constante.

Et alors $c = \frac{dx}{dt}$ on obtient encore une onde de concentration qui se propage à la vitesse c.

## II) Modèle d'écoulement du trafic.

### A- Première méthode d'ajustement.

Les séries statistiques qui ont servi à l'élaboration des modèles se composent de 240 couples de valeurs $(\tau, q)$ où $\tau$ est le taux d'occupation.

En un point P de la chaussée, $\tau$ est le temps pendant lequel il y a un véhicule sur P. En fait, si pendant le temps $\Delta t$, le point P est occupé par un véhicule pendant le temps $\Delta\theta$ sont taux d'occupation vaut :

$$\tau = \frac{\Delta\theta}{\Delta t}$$

Pour les valeurs voisines de $\tau$, on observe de grandes variations de q, ce qui fait admettre que q ne dépend pas uniquement de $\tau$ ; pour minimiser le caractère aléatoire de q, on procède à des groupages de points ; le groupage est fait par paquets de 20 points suivant les valeurs croissantes de $\tau$ chaque paquet est remplacé par son point moyen.

#### i) Modèle à deux paramètres.

Plusieurs séries de données ont été utilisées dans cette partie.
Les observations sont faites toutes les minutes, les paramètres mesurés sont :
Le débit ( en véhicules par minute) q, et le taux d'occupation $\tau$.

Le problème ici est d'ajuster des courbes théoriques à des nuages de points, valeurs de u $(u_1, u_2, ..., u_k)$, ayant une forme convexe.
Pour cela, on commence par ajuster au nuage une ligne polygonale.

En effet, soit $y_1, y_2, ..., y_k$ les ordonnées des sommets d'une ligne polygonale C joignant les points $(\tau_1, y_1), (\tau_2, y_2), ..., (\tau_k, y_k)$. On leur impose la condition suivante :

En appelant f la fonction ayant pour graphe C :

a) $y = f(\tau)$

b) C est convexe ( puisque le nuage est convexe )

On pose alors :

$$y_i = f(\tau_i) \qquad pour \;\; i = 1, 2, ..., k$$

La condition b) se traduit immédiatement par :